

Prepared in cooperation with the National Park Service

Monitoring and Real-time Modeling of *Escherichia coli* Bacteria for the Chattahoochee River, Chattahoochee River National Recreation Area, Georgia, 2000–2019



Open-File Report 2020–1048

Front cover. *Top*, Brent Aulenbach paddling at Rescue Rock Rapid, Chattahoochee River, Chattahoochee River National Recreation Area, Gwinnett and Forsyth Counties, Ga. *Left*, Chattahoochee River, Chattahoochee River National Recreation Area, at Jumping Rock, Cobb and Fulton Counties, Ga. *Right*, Jeff Riley fishing in the Chattahoochee River, Chattahoochee River National Recreation Area, below Powers Ferry Road, Cobb and Fulton Counties, Ga. Photographs by Alan Cressler, U.S. Geological Survey.

Back cover: *Top*, Lydia Hughes and Manuel Beers paddling in the Chattahoochee River, Chattahoochee River National Recreation Area, on the east side of Bowmans Island, Gwinnett County, Ga. *Bottom*, Chattahoochee River, Chattahoochee River National Recreation Area, at Jumping Rock, Cobb and Fulton Counties, Ga. Photographs by Alan Cressler, U.S. Geological Survey.

Monitoring and Real-time Modeling of *Escherichia coli* Bacteria for the Chattahoochee River, Chattahoochee River National Recreation Area, Georgia, 2000–2019

By Brent T. Aulenbach and Anna M. McKee

Prepared in cooperation with the National Park Service

Open-File Report 2020–1048

**U.S. Department of the Interior
U.S. Geological Survey**

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2020

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <https://store.usgs.gov/>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Aulenbach, B.T., and McKee, A.M., 2020, Monitoring and real-time modeling of *Escherichia coli* bacteria for the Chattahoochee River, Chattahoochee River National Recreation Area, Georgia, 2000–2019: U.S. Geological Survey Open-File Report 2020–1048, 43 p., <https://doi.org/10.3133/ofr20201048>.

ISSN 2331-1258 (online)

Acknowledgments

The Chattahoochee River BacteriALERT is a cooperative project with the National Parks Service, the Chattahoochee Parks Conservancy, the Chattahoochee Riverkeeper, the City of Roswell, Ga. Water Utility, the Cobb County Water System, and the U.S. Geological Survey. Thanks also goes to the Atlanta Regional Commission, Forsyth County, Ga., Georgia Power, Gwinnett County, Ga., and the U.S. Army Corps of Engineers for support of stream gages used for modeling within this report.

Contents

Acknowledgments	iii
Abstract	1
Introduction.....	1
Background.....	2
Purpose and Scope	4
Methods.....	5
Study Area.....	6
<i>Escherichia coli</i> (<i>E. coli</i>) Sampling and Analysis.....	6
Surface-Water Monitoring	7
Bacteria Modeling	7
Outlier Analysis and Model Assessment	13
Results	14
<i>Escherichia coli</i> (<i>E. coli</i>) Monitoring.....	14
Outlier Analysis	17
<i>Escherichia coli</i> (<i>E. coli</i>) Models.....	17
Model Assessment.....	24
Model Evaluation for Beach Action Value.....	33
Discussion.....	33
Evaluation of Model Explanatory Variables.....	33
Assessment of <i>Escherichia coli</i> (<i>E. coli</i>) Trend Explanatory Variables.....	34
Assessment of Outliers	35
Model Evaluation for Risk Assessment.....	35
Comparison with Previous Models	36
Summary and Conclusions.....	37
References Cited.....	38
Appendix 1. Outliers Removed from Regression Analysis	41

Figures

1. Map showing the Chattahoochee River National Recreation Area and watershed basin, and BacteriALERT program monitoring stations.....5
2. Number of *Escherichia coli* samples collected by station and year for the period October 23, 2000, to May 23, 2019.....10
3. Map showing the Chattahoochee River National Recreation Area and watershed basin, BacteriALERT program monitoring stations, and nine tributary watersheds monitored for streamflow
4. Sample *Escherichia coli* densities on the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 23, 2000, to May 23, 2019 (May 5, 2016, to May 23, 2019, for Powers Ferry)15

5. Scatterplots of <i>Escherichia coli</i> densities from samples collected on the same day between stations on the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 23, 2000, to May 23, 2019 (May 5, 2016 to May 23, 2019 for Powers Ferry)	16
6. Graphs showing <i>Escherichia coli</i> density versus <i>A</i> , time; <i>B</i> , season; <i>C</i> , water temperature; <i>D</i> , streamflow; <i>E</i> , upstream tributary streamflow; and <i>F</i> , turbidity at Chattahoochee River near Norcross, Ga. (USGS station 02335000) for the period October 1, 2007, to May 23, 2019	18
7. Graphs showing <i>Escherichia coli</i> density versus <i>A</i> , time; <i>B</i> , season; <i>C</i> , water temperature; <i>D</i> , streamflow; <i>E</i> , upstream tributary streamflow; and <i>F</i> , turbidity at Chattahoochee River at Powers Ferry and I-285 near Atlanta, Ga. (USGS station 02335880) for period June 16, 2016, to May 23, 2019	19
8. Scatterplots showing <i>Escherichia coli</i> density versus <i>A</i> , time; <i>B</i> , season; <i>C</i> , water temperature; <i>D</i> , streamflow; <i>E</i> , upstream tributary streamflow; and <i>F</i> , turbidity at Chattahoochee River at Atlanta, Ga. (USGS station 02336000) for the period October 4, 2007, to May 23, 2019	20
9. Comparison plots of 15-minute interval time-series versus calibration sample distributions of the model explanatory variables (turbidity, at-site streamflow, and upstream tributary streamflow) for the stations on the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 1, 2007, to May 30, 2019	21
10. Scatterplots of observed versus model predicted <i>Escherichia coli</i> densities for the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 4, 2007, to May 23, 2019 (June 16, 2016, to May 23, 2019, for Powers Ferry)	25
11. Scatterplots of <i>Escherichia coli</i> density model residuals versus model predicted densities for the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 4, 2007, to May 23, 2019 (June 16, 2016, to May 23, 2019, for Powers Ferry)	26
12. Scatterplots of <i>Escherichia coli</i> density model residuals versus <i>A</i> , time; <i>B</i> , season; <i>C</i> , water temperature; <i>D</i> , at-site streamflow; <i>E</i> , upstream tributary streamflow; and <i>F</i> , turbidity at Chattahoochee River near Norcross, Ga. (USGS station 02335000) for period October 1, 2007, to May 23, 2019	27
13. Scatterplots of <i>Escherichia coli</i> density model versus <i>A</i> , time; <i>B</i> , season; <i>C</i> , water temperature; <i>D</i> , at-site streamflow (from USGS station 02336000; lagged by –1 hour); <i>E</i> , upstream tributary streamflow; and <i>F</i> , turbidity at Chattahoochee River at Powers Ferry and I-285 near Atlanta, Ga. (USGS station 02335880) for period June 16, 2016, to May 23, 2019	28
14. Scatterplots of <i>Escherichia coli</i> density model residuals versus <i>A</i> , time; <i>B</i> , season; <i>C</i> , water temperature; <i>D</i> , at-site streamflow; <i>E</i> , upstream tributary streamflow; and <i>F</i> , turbidity at Chattahoochee River at Atlanta, Ga. (USGS station 02336000) for period October 4, 2007, to May 23, 2019	29
15. Histogram (<i>A</i>), outlier box plot (<i>B</i>), and normal quantile plot (<i>C</i>), for <i>Escherichia coli</i> density model residuals for the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 4, 2007, to May 23, 2019 (June 16, 2016, to May 23, 2019, for Powers Ferry)	30

16. Pie charts showing the frequency observed and model predicted <i>Escherichia coli</i> densities complied with or exceeded the Beach Action Value of 235 colony forming units per 100 mL at the three BacteriALERT stations for the model calibration period (October 4, 2007, to May 23, 2019; June 16, 2016, to May 23, 2019, for Powers Ferry)	31
17. Example of a time-series of <i>Escherichia coli</i> density model predictions and uncertainty for the Chattahoochee River at Norcross (USGS station 02335000) for the period August 7, 2013, to August 10, 2013	32

Tables

1. EPA recommended 2012 Recreational Water Quality Criteria and Beach Action Values for <i>E. coli</i> in fresh waters for primary contact recreation	3
2. Proposed <i>E. coli</i> indicator bacteria criteria for drinking water and non-estuarine fishing designated uses for Georgia	3
3. List of stations used in this analysis including locations, elevations, and cooperators	8
4. List of stations used in this analysis with watershed drainage area, location on Chattahoochee River, and travel and lag times from Buford Dam assuming a 2.6 miles per hour stream velocity	9
5. BacteriALERT <i>Escherichia coli</i> sampling through May 23, 2019, including dataset start date and statistical distribution of densities	14
6. Number of samples exceeding the Environmental Protection Agency's Beach Action Value of 235 cfu/100 mL <i>E. coli</i> , the geometric mean of 126 cfu/100 mL <i>E. coli</i> , and the statistical threshold value of 410 cfu/100 mL <i>E. coli</i> for the period October 23, 2000, to May 23, 2019 (May 5, 2016, to May 23, 2019, for Powers Ferry)	17
7. Summary of outliers removed from regression analysis, for the period October 4, 2007, to May 23, 2019 (June 16, 2016, to May 23, 2019, for Powers Ferry)	22
8. Summary of <i>Escherichia coli</i> model statistics and the Duan's smearing estimator	22
9. <i>Escherichia coli</i> model parameter coefficients.....	22
10. <i>Escherichia coli</i> model parameter estimates and statistics	23
11. <i>Escherichia coli</i> model confidence intervals calculated from the model root-mean-square error for ± 1 standard deviation (SD), 95-percent confidence intervals, and for the Beach Action Value of 235 colony forming units per 100 mL using ± 1 SD	24
12. Percentage of time predicted <i>Escherichia coli</i> densities that exceeded the Beach Action Value of 235 colony forming units per 100 mL at the three BacteriALERT stations	33
13. Comparison of <i>Escherichia coli</i> model statistics between models developed in this report and in Lawrence (2012)	36

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
Length		
mile (mi)	1.609	kilometer (km)
Area		
square mile (mi ²)	259.0	hectare (ha)
square mile (mi ²)	2.590	square kilometer (km ²)
Volume		
gallon (gal)	3785	milliliter (mL)
cubic inch (in ³)	16.39	milliliter (mL)
cubic foot (ft ³)	28320	milliliter (mL)
Flow rate		
cubic foot per second (ft ³ /s)	0.02832	cubic meter per second (m ³ /s)
mile per hour (mi/h)	1.609	kilometer per hour (km/h)

International System of Units to U.S. customary units

Multiply	By	To obtain
Length		
kilometer (km)	0.6214	mile (mi)
Area		
hectare (ha)	2.471	acre
square hectometer (hm ²)	2.471	acre
hectare (ha)	0.003861	square mile (mi ²)
square kilometer (km ²)	0.3861	square mile (mi ²)
Volume		
milliliter (mL)	0.0002642	gallon (gal)
milliliter (mL)	0.06102	cubic inch (in ³)
milliliter (mL)	0.00003531	cubic foot (ft ³)
Flow rate		
cubic meter per second (m ³ /s)	70.07	acre-foot per day (acre-ft/d)
cubic meter per second (m ³ /s)	35.31	cubic foot per second (ft ³ /s)
kilometer per hour (km/h)	0.6214	mile per hour (mi/h)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$^{\circ}\text{F} = (1.8 \times ^{\circ}\text{C}) + 32.$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$^{\circ}\text{C} = (^{\circ}\text{F} - 32) / 1.8.$$

Elevation, as used in this report, refers to distance above the vertical datum.

Supplemental Information

Densities of bacteria in water are given in either counts per 100 milliliters (counts/100 mL), colony forming units per 100 milliliters (cfu/100 mL), or most probable number per 100 milliliters (MPN/100 mL).

Abbreviations

AIC	Akaike Information Criteria
ASTM	American Society for Testing and Materials
BAV	U.S. Environmental Protection Agency's Beach Action Value
BIC	Bayesian Information Criteria
BMPs	Best Management Practices
°C	degrees Celsius
cfs	cubic feet per second
cfu	colony forming units
CRNRA	Chattahoochee River National Recreation Area
CSO	combined sewer overflow
<i>E. coli</i>	<i>Escherichia coli</i>
EPA	U.S. Environmental Protection Agency
FIB	fecal indicator bacteria
FNU	Formazin Nephelometric Units
GI	gastrointestinal
GM	geometric mean
mL	milliliters
MPN	most probable number
NEEAR	EPA's National Epidemiological and Environmental Assessment of Recreational Water
NEEAR-GI	National Epidemiological and Environmental Assessment of Recreational Water - gastrointestinal illness
NGI	National Epidemiological and Environmental Assessment of Recreational Water gastrointestinal illness
NWIS	National Water Information System
OLS	ordinary least squares (regression)
RMSE	root-mean-square error
RPD	relative percent difference
R ²	coefficient of determination
RWQC	EPA's 2012 Recreational Water Quality Criteria
SAWSC	USGS South Atlantic Water Science Center
SD	standard deviation
STV	statistical threshold value
USGS	U.S. Geological Survey

Monitoring and Real-time Modeling of *Escherichia coli* Bacteria for the Chattahoochee River, Chattahoochee River National Recreation Area, Georgia, 2000–2019

By Brent T. Aulenbach and Anna M. McKee

Abstract

The Chattahoochee River National Recreation Area (CRNRA; <https://www.nps.gov/chat/index.htm>) is a National Park Service unit/park with 48 miles of urban waterway in the Atlanta metropolitan area. The Chattahoochee River within the CRNRA is a popular place for water-based recreation but is known to periodically experience elevated levels of fecal-coliform bacteria associated with warm-blooded animals that can result in a variety of pathogen-related human illnesses. In 2000, the National Park Service entered into a public-private partnership with the U.S. Geological Survey (USGS) and the Chattahoochee Riverkeeper, called the Chattahoochee River BacteriALERT program, to monitor *Escherichia coli* (*E. coli*) which is a fecal indicator bacteria (FIB) and a proxy for human health risk from waterborne pathogens. The BacteriALERT network monitors *E. coli* densities at three stations on the Chattahoochee River within the CRNRA, at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000; <https://www2.usgs.gov/water/southeast/ga/bacteria/index.php>). *E. coli* densities determined from water samples were compared to the U.S. Environmental Protection Agency's Beach Action Value (BAV) of 235 colony forming units (cfu) per 100 milliliters (mL) to assess whether conditions were considered safe for freshwater, primary contact recreational use. Sample *E. coli* densities exceeded the BAV for 15.5 percent of the samples collected at Norcross ($n = 1,969$) and 30.3 percent of the samples at Atlanta ($n = 1,938$) for the study period October 23, 2000, to May 23, 2019, and 33.6 percent of the samples from Powers Ferry ($n = 134$) for the study period May 5, 2016, to May 23, 2019.

Models to predict *E. coli* densities in near real-time were developed for the three BacteriALERT stations. Models were developed using forward-stepwise multiple linear regression with the Bayesian Information Criteria and were calibrated with samples collected between October 4, 2007, and May 23, 2019. Explanatory variables included season, turbidity, water temperature, streamflow, upstream tributary streamflows, and temporal trend. The most statistically significant explanatory variables in the models were turbidity, upstream

tributary streamflows, and season. The Norcross model had an increasing trend in *E. coli* densities of 2.3 percent per year. A significant trend was not detected for the Atlanta station, while trends were not assessed for Powers Ferry models due to the short (3-year) calibration period. Model adjusted R^2 s ranged from 0.686 (Atlanta) to 0.795 (Norcross with time trend) indicating that the models explained a substantial portion of the variations in *E. coli* densities. Evaluation of model predictions and residuals indicated that models were well posed and exhibited little bias. The models performed well in accurately determining compliance and exceedance of the BAV with low misidentification rates ranging from 3.5 percent (Norcross) to 11.3 percent (Powers Ferry). Misidentification was most common for densities near the BAV, and misidentification rates in the study were low despite fairly low model precisions because *E. coli* densities were infrequently near the BAV. The precisions of the models developed herein were comparable to the more complex models developed by Lawrence (2012) that were never implemented in the BacteriALERT program due to their computational complexity. The predictive *E. coli* models developed herein will improve the ability to assess the health risks of water-based recreational activities in the CRNRA in near real-time.

Introduction

The Chattahoochee River National Recreation Area (CRNRA; <https://www.nps.gov/chat/index.htm>) is a National Park Service unit/park with 48 miles of urban waterway in the Atlanta metropolitan area, Georgia. In 2018, the CRNRA had 2,873,866 visitors and economic contributions of about 128 million dollars in visitor spending and 179 million dollars in economic output (Cullinane and others; 2019). Approximately 30 percent of the park visitation involves water-based recreation, including primary contact activities such as swimming and tubing, and secondary contact activities including canoeing, kayaking, boating, rowing, wading, and fishing. The CRNRA portion of the Chattahoochee River is the southern-most designated trout fishery with a reproducing population in North America (National Park Service, 2011)

with a reproducing population of brown trout and is periodically stocked with rainbow trout. However, the lower reach of the CRNRA section of the Chattahoochee River frequently exceeds fecal indicator bacteria (FIB) standards for designated uses of the river for recreation and drinking water (Gregory and Frick, 2000; Hartel and others, 2004; Lawrence, 2012). A recent microbial source tracking study within the CRNRA watershed indicated that humans and dogs were the primary contributors of bacterial contamination among sources tested (humans, dogs, and ruminants; McKee and others, 2020). Therefore, fecal contamination from point and nonpoint-source runoff is a major health concern for park management. In 2000, the National Park Service entered into a public-private partnership with the U.S. Geological Survey (USGS) and the Chattahoochee Riverkeeper, called the Chattahoochee River BacteriALERT program, to monitor *Escherichia coli* (*E. coli*) densities within the CRNRA as a proxy for human health risk from waterborne pathogens (Leclerc and others, 2001; Tallon and others, 2005). Pathogens are disease causing bacteria, viruses, and other microorganisms. This report summarizes *E. coli* monitoring and documents regression models developed to make near-real-time estimates for *E. coli* densities at three locations on the Chattahoochee River that are encompassed by the CRNRA. The *E. coli* estimates can be used to inform water-based recreational users of the risk from waterborne pathogens and will be publicly available on the USGS BacteriALERT website (<https://www2.usgs.gov/water/southeastlantic/ga/bacteria/index.php>). The website received over 39,000 visits in 2018 (Richard “Scott” Young, USGS, written commun., August 14, 2019).

Background

Water-based recreation in rivers, streams, lakes, reservoirs, estuaries, and marine water bodies with fecal contamination from warm-blooded animals can result in a variety of pathogen-related human illnesses. Human sources of fecal-associated bacteria in surface waters include leaking septic tanks, sewer overflows, conveyance leaks, and from perturbations at sewage treatment plants. Animal sources in surface waters include direct inputs and storm runoff of fecal matter from wildlife, livestock, and dogs (Soller and others, 2010; Riedel and others, 2015). Gastrointestinal (GI) illnesses can develop within 10 to 12 days after contact with contaminated waters. EPA’s National Epidemiological and Environmental Assessment of Recreational Water (NEEAR; Wade and others, 2010) describes the common symptoms of this GI illness to include diarrhea, vomiting, and nausea and stomachache (illness definition referred to as NEEAR-GI or NGI). Some illnesses can be serious with possible fatal complications.

The human health risk from water-borne pathogens is usually assessed from densities of FIB such as *E. coli*, *Enterococci*, and fecal coliform. Fecal indicator bacteria originate from the same sources as the pathogens of interest but do not always accurately reflect the human risk attributable to

pathogens. Furthermore, the human health risk may be dependent on the source of the fecal contamination, with initial studies indicating that FIBs from human and fresh cattle feces are associated with similar illness rates while FIBs from fresh gull, chicken, and pig feces are associated with lower illness rates than human sources (Schoen and Ashbolt, 2010; Soller and others, 2010). While *E. coli* is not known to replicate outside of the gastrointestinal tracts of warm-blooded animals, other types of fecal coliforms have been shown to replicate in natural environments (McLellan and others, 2001). Fecal indicator bacteria can survive for several days in surface waters and for months in lake sediments where they are protected from viral or bacterial predators and from ultraviolet light that can inactivate bacteria such that they cannot replicate and cause infection (Darakas, 2002). Fecal indicator bacteria are used for regulatory purposes because they can be determined in a cost effective and timely manner, whereas monitoring for pathogens can require separate analyses for each individual pathogen and can be expensive and time consuming.

The reach of the Chattahoochee River that encompasses the CRNRA has designated uses of recreation and drinking water (Buford Dam to Atlanta at inflow of Peachtree Creek; Rule 391-3-6-.03 Water Use Classifications and Water Quality Standards, effective August 16, 2016). As part of the Clean Water Act, states are required to develop water-quality standards that are protective of a water-body’s designated environmental use. Georgia’s current and proposed water-quality standards using *E. coli* as the FIB are summarized for recreation use in table 1 and for drinking water use in table 2 (proposed). Further details on these standards and their implementation are provided in the sidebar “Georgia Water Quality Standards for Recreation and Drinking Water Use Designations.”

In addition to the designated use water-quality standards, the EPA has developed the Beach Action Value (BAV) as a precautionary tool for making beach notification advisories (table 1; U.S. Environmental Protection Agency, 2012). The BAV criteria uses a single sample, thereby providing timely notification of elevated levels of FIB for recreational advisories. The BAV represents to the 75th percentile value of a water-quality distribution used to develop the Recreational Water Quality Criteria (RWQC) for its corresponding illness rate.

The Chattahoochee River BacteriALERT program monitors *E. coli* densities at three locations on the Chattahoochee River covering a 28-mile stretch of the lower portion of the CRNRA (fig. 1). Monitoring started in October 2000 at Norcross and Atlanta (USGS stations 02335000 and 02336000, respectively) and a third station was added in May 2016 at Powers Ferry (02335880). The Powers Ferry station was added to improve the spatial resolution of bacteria densities in the lower section of the park below Morgan Falls Dam where the majority of the recreation in the CRNRA occurs. Recreation occurs particularly between Powers Island, which is adjacent to the Powers Ferry station, and Paces Mill, which is less than one mile upstream of the Atlanta station.

Table 1. EPA recommended 2012 Recreational Water Quality Criteria (RWQC) and Beach Action Values (BAVs) for *E. coli* in fresh waters for primary contact recreation.

[Georgia has adopted the RWQC for the estimated NGI illness rate of 36 per 1,000 primary contact recreators. Abbreviations: NGI, National Epidemiological and Environmental Assessment of Recreational Water gastrointestinal illness; cfu/100 mL, colony forming units per 100 milliliters; EPA, Environmental Protection Agency]

Estimated NGI illness rate (per 1,000 primary contact recreators)	Geometric mean (cfu/100 mL)	Statistical threshold value (cfu/100 mL)	Beach Action Value (cfu/100 mL)
36	126	410	235
32	100	320	190

Table 2. Proposed *E. coli* indicator bacteria criteria for drinking water and non-estuarine fishing designated uses for Georgia (Booth and Adams, 2018).

[Abbreviations: Oct., October; Non-human GM, geometric mean criteria for variance when non-human *E. coli* source does not exceed 126 counts/100 mL; Nov., November; mL, milliliters; —, criteria not defined]

Season	Geometric mean (counts/100 mL)	Statistical threshold value (counts/100 mL)	Non-human GM (lakes and reservoirs; counts/100 mL)	Non-human GM (free flowing freshwater streams; counts/100 mL)
Recreation (May–Oct.)	126	410	189	315
Non-recreation (Nov.–April)	630	2050	—	—

Two of the four miles of the river between Powers Island and Paces Mill are adjacent to the Atlanta city limits, where heavier bacterial inputs to the Chattahoochee River associated with dense urban development might be expected. Bacteria densities at the upstream Norcross station are not expected to be representative of densities near the Powers Ferry station because the Norcross station is 24.5 miles upstream and the Chattahoochee River flows through intervening Bull Sluice Lake and Morgan Falls Dam, which can affect water quality by reducing transport of sediment-related bacteria and promote inactivation of bacteria by increasing exposure to ultraviolet light within the shallow lake.

Streamwater samples were collected at each of the stations about once per week and analyzed for *E. coli*. Results of the analysis were then posted to the publicly available USGS BacteriALERT website. When *E. coli* densities exceeded the EPA designated BAV of 235 cfu per 100 mL (corresponding to an estimated NGI illness rate of 36 per 1,000 primary contact recreators, [table 1](#)), a health advisory was posted for that location, the Park canceled all Park-led water-based recreational activities at that location, and concessionaires (for example, rafting, kayaking, and tubing outfitters) utilizing the Park were required to notify their clients of possibly unsafe water-quality conditions but were still allowed to operate.

Due to the time required to process and incubate the samples, *E. coli* results are not available until about 18+ hours after sampling. Furthermore, short-term increases in bacteria levels associated with stormflow and sanitary sewer overflow events (Gregory and Frick, 2000) are often not representative of the densities observed in the weekly samples. To provide

more timely water-quality advisories to recreational users, the USGS developed predictive models of *E. coli* densities on the Chattahoochee River at the Norcross and Atlanta stations (Lawrence, 2012). While Lawrence (2012) developed a set of models with varying degrees of complexity for each station, the simplest models were implemented to predict *E. coli* densities in near real-time. These simple models used only turbidity as a water-quality surrogate for *E. coli* densities as its only explanatory variable, which was measured from water-quality sondes installed at these stations. Turbidity is a measure of how cloudy or opaque water is and is the result of the amount and type of suspended and dissolved matter in the water (for example, clay, silt, finely divided organic matter, plankton and other microscopic organisms, organic acids, and dyes; American Society for Testing and Materials International, 2003; Anderson, 2005) and can be affected by the color of the water. Bacteria densities tend to be higher when turbidities are higher due to the association of bacteria with suspended sediments. Studies have reported that 34 to 42 percent of *E. coli* were attached to suspended sediments in surface-water samples (Fries and others, 2006; Krometis and others, 2007). Currently (as of 2020), *E. coli* estimates from these models are available on the BacteriALERT website in near-real-time as data is generally transmitted and estimates are updated hourly.

Lawrence (2012) did an extensive analysis of *E. coli* densities on the Chattahoochee River at the Norcross and Atlanta stations. This included improving *E. coli* model predictions by including additional explanatory variables other than turbidity. That study indicated that higher *E. coli* densities were observed when there were increases in contributions

Georgia Water Quality Standards for Recreation and Drinking Water Use Designations

In 2012, the U.S. Environmental Protection Agency (EPA) updated its recommended FIB criteria for fresh and marine waters designated for recreational use (2012 Recreational Water Quality Criteria [RWQC], U.S. Environmental Protection Agency, 2012). The fresh waterbody *E. coli* criteria are summarized in [table 1](#). The State of Georgia, in their 2013 triennial review of water-quality standards, has established bacterial criteria for primary contact recreational waters using *E. coli* and *Enterococci* as FIBs (U.S. Environmental Protection Agency, 2016; Rule 391-3-6-.03(12)) and has adopted the rule for recreational waters other than coastal waters based on the EPA 2012 RWQC recommendations for *E. coli* for the estimated NGI illness rate of 36 per 1,000 primary contact recreators ([table 1](#); Rule 391-3-6-.03(6)(b)(i)(2)). Georgia has proposed changing the FIB criteria from fecal coliform to *E. coli* and *Enterococci* for drinking water and fishing (secondary recreational contact) designated waterbody uses in their 2016 triennial review of water quality standards (Booth and Adams, 2018). The proposed guidelines for drinking-water supply and non-estuarine fishing designations are stated in [table 2](#). These designated uses do not require as strict water-quality criteria as primary contact recreation.

The Georgia Environmental Protection Division periodically assesses whether waters with designated uses meet the water-quality criteria. Compliance is based on a minimum of four samples over a 30-day period collected at intervals not less than 24 hours and consists of two FIB criteria: (1) the geometric mean (GM) of the sample concentrations should not exceed the selected criteria's GM, and (2) no more than 10 percent of the sample concentrations should exceed the selected criteria's statistical threshold value (STV). The EPA RWQC recommendations were developed for the general population including children and have not been evaluated for persons over 55 years of age, pregnant women, or other vulnerable individuals such as those that are immune-compromised. The EPA indicates that the public should be advised of potential additional risk from sources of urban runoff and combined sewer overflows (CSOs) when compliance sampling is not done during or immediately after a rain event.

of stormflows from tributaries downstream of Buford Dam ([fig. 1](#)), which typically have higher bacteria densities. Bacteria densities in water discharged from Buford Dam were typically low (Gregory and Frick, 2000) as the result of water being released from near the bottom of the water column of Lake Lanier through the power generation turbines, which was cold and low in particulates. Higher flows coming out of Buford Dam are still expected to result in higher downstream bacteria densities due to flow-related increases in suspension of bacteria associated sediment but not as much of an increase as seen with higher flows from tributaries. These complex relations explain why at-site streamflow was found to be a poor explanatory variable for predicting *E. coli* densities at these two stations. Lawrence (2012) addressed the differences in the sources of streamflow in the *E. coli* density models by including streamflow regime and streamwater condition indicator variables to categorize streamflow sources and conditions. Additional details on these models are provided in the sidebar “Explanatory Variables Used in *E. coli* Density Modeling from Lawrence (2012).” While these variables improved *E. coli* density predictions, they required complex queries that were difficult to implement within the USGS National Water Information System (NWIS) real-time data framework. Therefore, BacteriALERT *E. coli* density real-time estimates employed the simplest models with turbidity as the sole explanatory variable.

Since the BacteriALERT program started in 2000, many tributaries between Buford Dam and the Chattahoochee River at Atlanta station have been monitored for streamflow as part of the City of Roswell, Cobb County Water System, Forsyth County, and Gwinnett County monitoring programs. As of 2007, there were nine tributaries that were monitored along this reach of the Chattahoochee River, representing about half of the tributary drainage areas along this reach. Flow data from these gages allows for a more direct way to model the effects of the contribution of tributary flows on *E. coli* densities of the river, recognizing that stormflow from these tributaries often have elevated *E. coli* densities.

Purpose and Scope

Given the potential human health risk to visitors of the CRNRA from water contact with fecal bacteria and the possible repercussions on Park recreational activities from overestimating bacteria levels, it is important to provide as accurate and timely advisories of the potential health risk of fecal bacteria as possible. Two recent steps were taken to expand and enhance the BacteriALERT monitoring program to improve *E. coli* advisories: (1) A third monitoring station

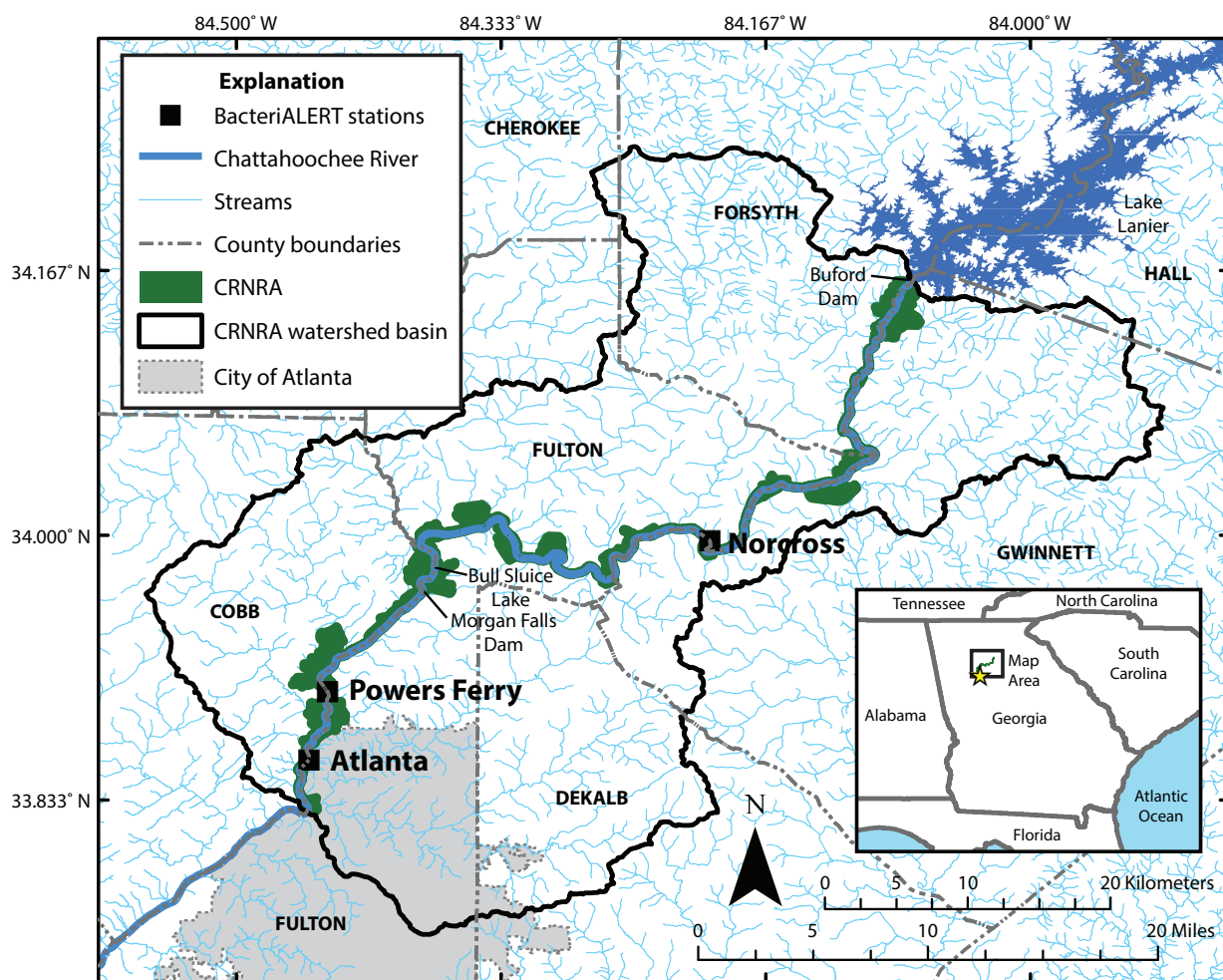


Figure 1. Map showing the Chattahoochee River National Recreation Area (CRNRA) and watershed basin, and BacteriALERT program monitoring stations. BacteriALERT program monitoring stations: Norcross (USGS station 02335000); Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000). The yellow star in the inset is the location of the City of Atlanta.

was added in May 2016 at Powers Ferry, located between the previously existing stations at Norcross and Atlanta, which improves the resolution of bacteria levels along the lower section of the CRNRA where the majority of recreation occurs. (2) The bacteria models for estimating *E. coli* for providing near real-time potential health risk advisories were updated and improved for the Norcross and Atlanta stations and developed for the new Powers Ferry station using tributary flow explanatory variables such that the models better capture source contributions of bacteria. This report assesses *E. coli* density monitoring data for the three BacteriALERT monitoring stations and documents newly developed regression models for predicting *E. coli* densities at these stations in real-time to be used for health risk advisories for primary recreational use in the CRNRA.

Methods

As part of the BacteriALERT program, *E. coli* densities have been monitored at three stations on the Chattahoochee River to inform recreators within CRNRA of potential health risks. Streamflow and continuous water-quality parameters have been monitored at these three stations, just downstream of Buford Dam, and along nine tributaries, and were related to variations in monitored *E. coli* densities. Predictive models of *E. coli* densities were developed for the three bacteria monitoring stations, using streamflow and water-quality data from these stations and lagged streamflows from nine upstream tributaries as explanatory variables, in order to provide near real-time assessment of health risks of water-based recreation.

Study Area

The CRNRA is a unique 48-mile stretch of urban waterway in the greater Atlanta metro area and accounts for over 60 percent of the green space in the greater Atlanta, Georgia, area (National Park Service, 2011; [fig. 1](#)). This small, naturally shallow reach of the Chattahoochee River supplies the majority of water for millions of people in the Atlanta metropolitan area, assimilates wastewater from municipal wastewater treatment plants, and receives stormwater from tributaries draining urban areas. The Chattahoochee River designated water use from Buford Dam to Atlanta (to inflow of Peachtree Creek) is classified as recreation and drinking water. Chattahoochee River flows within the CRNRA are controlled by regulated releases from Lake Lanier at Buford Dam and Bull Sluice Lake at Morgan Falls Dam ([fig. 1](#)). Lawrence (2012) provides a more detailed description of the study area and its climate, streamflow characterization, dam operations, and previous bacteria studies in the study area.

Escherichia coli (*E. coli*) Sampling and Analysis

As part of the BacteriALERT program, *E. coli* densities were monitored at three stations on the Chattahoochee River over a 27.8-mile stretch within the CRNRA. These stations include (1) Norcross, the most upstream station, located at Medlock Bridge Road (Georgia State Highway 141) 17.75 miles downstream of Buford Dam; (2) Powers Ferry, located at Powers Ferry Road NW and Interstate 285, about 6.5 miles downstream of Morgan Falls Dam; and (3) Atlanta, the most downstream station, located at Paces Ferry Road near the lower end of the CRNRA ([fig. 1](#), [tables 3](#) and [4](#)). Bacteria monitoring at the Norcross and Atlanta stations started in October 2000 and the Powers Ferry station started in May 2016. Samples were collected at various sampling frequencies and more recently were collected weekly ([fig. 2](#)). Employees and volunteers of the Chattahoochee Riverkeeper and CRNRA collected samples off bridges at midchannel as a single vertically-integrated sample to improve the sample representation of the river over a grab sample as outlined in the National Field Manual for the Collection of Water-Quality Data (U.S. Geological Survey, various dates).

Samples were processed and analyzed by CRNRA personnel at the USGS South Atlantic Water Science Center (SAWSC) microbiology laboratory in Norcross, Georgia. *E. coli* bacteria densities were determined using the Colilert®-18 medium and Quanti-Tray® system manufactured by the IDEXX Corporation (IDEXX Laboratories, Inc., 2002a, b). Analytical procedures followed standard method 9223B (Clesceri and others, 1998), which is approved by the EPA for ambient water analyses (U.S. Environmental Protection Agency, 2003). In this method, the sample is sealed into a tray with multiple discrete wells with the Colilert®-18 medium, which contains a growth medium and a fluorogen enzyme substrate. Samples are then incubated for 18 to

Explanatory Variables Used in *E. coli* Density Modeling from Lawrence (2012)

In addition to simplest models that only used turbidity as a surrogate for *E. coli* densities, Lawrence (2012) developed additional sets of regression models that included several additional indicator explanatory variables to improve model predictions. Indicator variables are categorical variables and only apply to the model when certain conditions are met. Two indicator variables were included to improve the modeling of differences in the sources of streamflow: a streamflow regime “EVENT” variable, and a streamwater condition “HCOND” variable. The EVENT variable categorizes streamflow conditions into (1) dry-weather flow, which represented water releases from Buford Dam or Morgan Falls Dam and (2) stormflow, which represented increases in the stream stage of the tributaries. Stormflow conditions required recent precipitation (0.75 inches of rain in the previous 48 hours) and increases in tributary flows (Suwanee Creek for the Norcross station; Rottenwood or Sope Creek for the Atlanta station) to avoid false identification from streamflow from dam releases. The streamwater condition “HCOND” variable was defined to categorize streamflow conditions into six categories: (a) StableLow, (b) StableNorm, and (c) StableHigh represented conditions where streamflow is relatively stable (controlled by water releases from the dams) and stage was low, normal, or high, respectively, while (d) RisingQ, (e) FallingQ, and (f) PeakQ represented various stormflow hydrograph conditions, presumably when tributaries were contributing substantial stormflow to Chattahoochee River streamflow. The three stormflow hydrograph HCOND categories were included to address the relation in which suspended sediment and related bacteria densities are typically higher on the rising limb of the hydrograph and peaks before peak streamflow. Low, normal, and high stage ranges for stable conditions were defined separately for the Norcross and Atlanta stations while RisingQ and FallingQ conditions were defined as increases or decreases in stream stage changes of greater than 5 percent per hour, respectively.

Water temperature and a seasonal indicator variable with warm and cool seasons were included in the models to address that bacteria densities are generally higher during the summer when water temperatures are higher. All explanatory variables mentioned were used in at least one of the models developed by Lawrence (2012) for stations at Norcross and Atlanta.

The indicator variables were used to fit separate offsets to the *E. coli* model densities for each condition. In the case of the HCOND indicator variable, a single

coefficient was fit to six conditions that are assigned sequential integer values which implies that the effect of streamflow condition varied by the order and quantity of their assigned indicator value and such that the incremental effects between each condition are the same, which might not be realistic. In some models an interaction variable was used as a combination of the EVENT and turbidity variables which allow for separate relations between *E. coli* densities and turbidity for dry-weather and stormflow streamflow conditions.

22 hours at $35 \pm 0.5^\circ\text{C}$. The fluorogen reacts with the enzymes produced by *E. coli*, which causes the medium to fluoresce under a long-wave ultraviolet light (366 nanometers [nm]), indicating the presence of one or more *E. coli* bacteria colonies in a well. The Colilert®-18 medium also selectively suppresses other non-coliforms that may interfere with *E. coli* bacteria growth or produce false positives. This method is a most probable number (MPN) technique in which the most likely density of bacteria is statistically determined from the number of positive wells. In this study, Quanti-Tray®/2000 trays were used, which divides the sample into 49 large and 48 small wells and can estimate the MPN of colonies from 1 to 2,419 without dilution. Typically, samples were analyzed using two different dilutions to allow for higher range of densities of bacteria to be measured without knowing the density of bacteria beforehand. Sample *E. coli* densities are available to the public from the USGS NWIS web interface (<https://waterdata.usgs.gov/nwis/>; water quality data category, field/lab samples data type, *E. coli* parameter code 50468, using USGS station numbers in table 3).

Surface-Water Monitoring

Stream surface water elevation (stage) and streamflow were determined on the Chattahoochee River at the three bacteria monitoring stations and at a station just downstream of Buford Dam, and at nine stations on tributaries of watersheds within the CRNRA watershed basin (fig. 3, tables 3 and 4). Streamflows at the BacteriALERT stations and tributaries were used as explanatory variables in the *E. coli* models while streamflow from the station downstream of Buford Dam was used along with streamflows from the BacteriALERT stations to determine downstream travel times on the Chattahoochee River. Stream stage was measured every 15 minutes and streamflow was computed at all stations except for the Chattahoochee River at Powers Ferry station where a stage-discharge relation had not been developed at the time of this study. Tributary flows were monitored as part of USGS cooperative programs with the City of Roswell, Cobb County Water System, Forsyth County, and Gwinnett County. Streamflows were determined by the SAWSC using standard

USGS protocols for measuring stage, making streamflow measurements, and computing discharge (Rantz and others, 1982a, b).

Continuous water-quality monitors that measured water temperature, turbidity, and specific conductance at 15-minute intervals were deployed in 2002 at the Norcross and Atlanta stations, and in 2016 for the Powers Ferry station. Water-quality monitoring data was available at most of the other stations but was not used in this analysis. Water-quality monitors typically were cleaned and their calibration checked every 2 weeks; more frequently following hydrologic events or after observing abnormal readings, which could have been associated with fouling, instrument failure, or point sources or nonpoint sources of pollution. Water-quality monitors were maintained and their corresponding sensor records checked using the quality-assurance and quality-control procedures outlined in Wagner and others (2006).

All continuously monitored data (stage, streamflow, and water quality) were transmitted hourly by way of satellite communication and are available to the public from the USGS NWIS web interface (<https://waterdata.usgs.gov/nwis/>; surface water data category, historical observations data type, using USGS station numbers in table 3) as values and time-series plots, which also can be accessed at https://waterdata.usgs.gov/ga/nwis/current/?type=flow&group_key=basin_cd.

Bacteria Modeling

E. coli densities were predicted in near real-time using multiple linear regression models that estimate sample densities as a function of model parameters derived from continuously measured streamflow and water-quality variables from USGS monitoring stations. Modeling procedures follow USGS guidelines presented in Rasmussen and others (2009). Modeling approaches herein are informed from Lawrence (2012) and designed to allow the models to be operationalized. Variables in the models included: at-site streamflow, turbidity, water temperature, upstream tributary streamflow, day of year, and decimal year. Bacteria densities tend to be higher when at-site streamflow, upstream tributary streamflow, and at-site turbidities are higher, at least in-part as a result of resuspension of bacteria in the water column and associated increases in suspended sediment for which a substantial portion of the bacteria can be attached to (Fries and others, 2006; Krometis and others, 2007). Turbidity, which is the cloudiness of water due to the suspension of particles, is a water-quality surrogate for suspended sediment and can be a better indicator of high bacteria densities than streamflow due to its specific relation with suspended sediment. While streamflow typically is correlated with suspended sediment concentrations, they do not necessarily vary directly with each other, because suspended sediment concentrations are known to have a hysteretic response in many streams with higher values observed during

8 Monitoring and Real-time Modeling of *E. coli* Bacteria for the Chattahoochee River, CRNRA, Georgia, 2000–2019

Table 3. List of stations used in this analysis including locations, elevations, and cooperators.

[The three BacteriALERT monitoring stations on the Chattahoochee River are in bold. Abbreviations: Cobb County, Cobb County Water System (Ga.); Ga., Georgia; Ga. 120, Georgia State Route 120; N, North; na, not available; NPS, National Park Service; USACE, U.S. Army Corps of Engineers (Mobile District)]

USGS station number	Station name	Latitude	Longitude	Horizontal Datum	Elevation (feet)	Vertical datum	Cooperator(s)
02334430	Chattahoochee River at Buford Dam, near Buford, Ga.	34° 09' 25"	84° 04' 44"	NAD83	912.1	NAVD88	USACE
02334480	Richland Creek at Suwanee Dam Road, near Buford, Ga.	34° 07' 57"	84° 04' 12"	NAD27	924	NAVD88	Gwinnett County, Ga.
02334578	Level Creek at Suwanee Dam Road, near Suwanee, Ga.	34° 05' 47"	84° 04' 47"	NAD27	956.4	NAVD88	Gwinnett County, Ga.
02334620	Dick Creek at Old Atlanta Road, near Suwanee, Ga.	34° 04' 17"	84° 07' 49"	NAD27	909.4	NAVD88	Forsyth County, Ga.
02334885	Suwanee Creek at Suwanee, Ga.	34° 01' 56"	84° 05' 22"	NAD27	909.9	NAVD88	Gwinnett County, Ga.
02335000	Chattahoochee River near Norcross, Ga.	33° 59' 50"	84° 12' 07"	NAD83	878.2	NAVD88	USACE, Georgia Power, Cobb County, NPS, and Chattahoochee Riverkeeper ¹
02335350	Crooked Creek near Norcross, Ga.	33° 57' 54"	84° 15' 54"	NAD27	869.8	NAVD88	Gwinnett County, Ga.
02335757	Big Creek below Hog Wallow Creek at Roswell, Ga.	34° 01' 03"	84° 21' 12"	NAD83	940	NGVD29	City of Roswell, Ga., Water Utility
02335790	Willeo Creek at Ga. 120, near Roswell, Ga.	34° 00' 10"	84° 23' 40"	NAD27	856.2	NAVD88	Cobb County
02335870	Sope Creek near Marietta, Ga.	33° 57' 14"	84° 26' 36"	NAD83	881.4	NAVD88	Atlanta Regional Commission ¹
02335880	Chattahoochee River at Powers Ferry and I-285 near Atlanta, Ga.	33° 54' 08"	84° 26' 30"	NAD27	na	na	The Chattahoochee Parks Conservancy and Cobb County
02335910	Rottenwood Creek at Interstate N Parkway, near Smyrna, Ga.	33° 53' 37"	84° 27' 28"	NAD27	820.2	NAVD88	Cobb County
02336000	Chattahoochee River at Atlanta, Ga.	33° 51' 33"	84° 27' 16"	NAD83	750.3	NAVD88	USACE, Georgia Power, Cobb County, NPS, and Chattahoochee Riverkeeper ¹

¹A U.S. Geological Survey Federal Priorities Streamgage.

Table 4. List of stations used in this analysis with watershed drainage area, location on Chattahoochee River, and travel and lag times from Buford Dam assuming a 2.6 miles per hour stream velocity.

[The three BacteriALERT monitoring stations on the Chattahoochee River are in bold. Distance downstream of Buford Dam for tributaries represents their confluence with the Chattahoochee River as tributary stations are located somewhat upstream. Ga., Georgia; Ga. 120, Georgia State Route 120; N, North; na, not applicable]

USGS station number	Station name	Drainage area (square miles)	River mile on Chattahoochee River	Distance downstream of Buford Dam (miles)	Travel time (hours)	Lag time to Norcross station (hours)	Lag time to Powers Ferry station (hours)	Lag time to Atlanta station (hours)
02334430	Chattahoochee River at Buford Dam, near Buford, Ga.	1,040	348	0.5	0.2	7	16.5	17.5
02334480	Richland Creek at Suwanee Dam Road, near Buford, Ga.	9.37	346.6	1.9	0.7	6	15.5	16.5
02334578	Level Creek at Suwanee Dam Road, near Suwanee, Ga.	5.06	342.1	6.4	2.5	4.5	14	15
02334620	Dick Creek at Old Atlanta Road, near Suwanee, Ga.	6.9	341.2	7.3	2.8	4	13.5	14.5
02334885	Suwanee Creek at Suwanee, Ga.	47.1	338	10.5	4.1	3	12.5	13.5
02335000	Chattahoochee River near Norcross, Ga.	1,170	330.75	17.75	6.9	0	na	na
02335350	Crooked Creek near Norcross, Ga.	8.87	325	23.5	9.1	na	7.5	8.5
02335757	Big Creek below Hog Wallow Creek at Roswell, Ga.	103	317.3	31.2	12.0	na	4.5	5.5
02335790	Willeo Creek at Ga. 120, near Roswell, Ga.	16.1	315	33.5	12.9	na	3.5	4.5
02335870	Sope Creek near Marietta, Ga.	30.7	308.6	39.9	15.4	na	1	2
02335880	Chattahoochee River at Powers Ferry and I-285 near Atlanta, Ga.	1,420	306.3	42.2	16.3	na	0	na
02335910	Rottenwood Creek at Interstate N Parkway, near Smyrna, Ga.	18.6	304.3	44.2	17.1	na	−0.5	0.5
02336000	Chattahoochee River at Atlanta, Ga.	1,450	303	45.5	17.6	na	−1	0

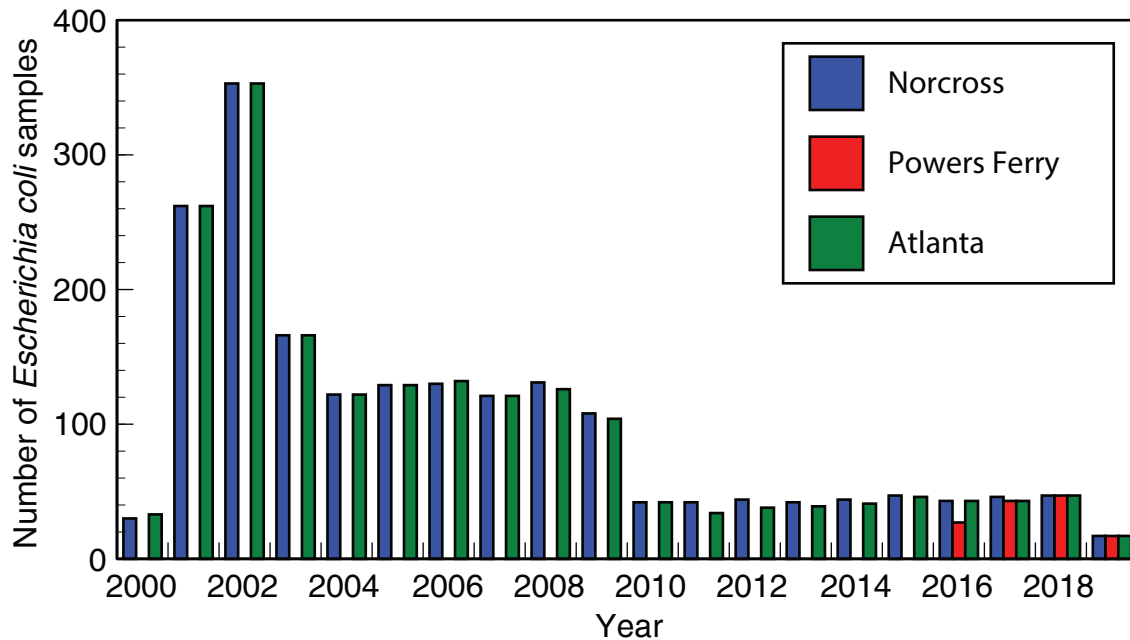


Figure 2. Number of *Escherichia coli* (*E. coli*) samples collected by station and year for the period October 23, 2000, to May 23, 2019. Norcross, USGS station 02335000; Powers Ferry, USGS station 02335880; Atlanta, USGS station 02336000.

the rising limb of a storm hydrograph (Seeger and others, 2004). Water temperature and day of year variables were used to model seasonal variability, whereby bacteria densities and water temperature were higher in the Chattahoochee River during the summer months. Darakas (2002) indicated that *E. coli* released into the environment can survive up to several days. Survival times were a positive function of the duration of its maintenance phase, which was found to peak in duration at 10 °C but remained high at 20 °C. These seasonal variables were intended to improve upon the warm/cool seasonal indicator variable used in Lawrence (2012) by allowing for additional temporal flexibility in the seasonal relation. The decimal year variable was used to model any long-term trends.

The purpose of including both at-site and upstream tributary flows was to discern the proportion of flow coming from Buford Dam versus from tributary watersheds within the CRNRA basin. The water released through Buford Dam is typically from near the bottom of the water column of Lake Lanier, which is expected to have relatively low *E. coli* densities due to its cold temperature and long residence time. Gregory and Frick (2000) indicated that median fecal-coliform bacteria densities (an FIB related to *E. coli*) in waters released through the dam were less than 20 colonies per 100 mL. While higher bacteria densities are expected when flows are higher from Buford Dam due to suspension of bacteria in the downstream river channel, even higher densities have been observed when the source of streamflow is from stormflows in the tributary watersheds. At-site streamflow was used in the models to represent variability in flows from Buford Dam. While this variable encompasses both the variability in flows

from Buford Dam and the tributaries, when used in concert with the upstream tributary flow variable, it models just the variability in flows from Buford Dam because the upstream tributary flow variable effectively captures the portion of variability attributed to the tributary flows. Lawrence (2012) modeled for these variations in sources of bacteria using the EVENT and HCOND indicator variables (see sidebar, “Explanatory Variables Used in *E. coli* Density Modeling from Lawrence (2012)”) to indicate whether tributary stormflows contributed to overall streamflows.

Since the analysis by Lawrence (2012), several additional tributary streamflow monitoring stations have been added as part of several USGS urban studies programs (table 3), providing a more comprehensive monitoring of tributary watersheds within the CRNRA basin (fig. 3) that allows for tributary inflows to be used as variables in the *E. coli* density models developed herein. As of October 1, 2007, nine tributaries have been consistently monitored within the CRNRA basin upstream of the BacteriALERT stations. Four are upstream of the Norcross station, eight are upstream of the Powers Ferry station, and nine are upstream of the Atlanta stations. The monitored tributaries represent 53, 60, and 60 percent of the CRNRA basin drainage areas below Buford Dam and BacteriALERT stations at Norcross, Powers Ferry, and Atlanta, respectively. The portion of the drainage areas should sufficiently capture the variability of tributary streamflow contributions necessary for modeling because intervening adjacent ungaged drainage areas should have reasonably similar stormflow responses as they would be expected to receive similar amounts of precipitation to

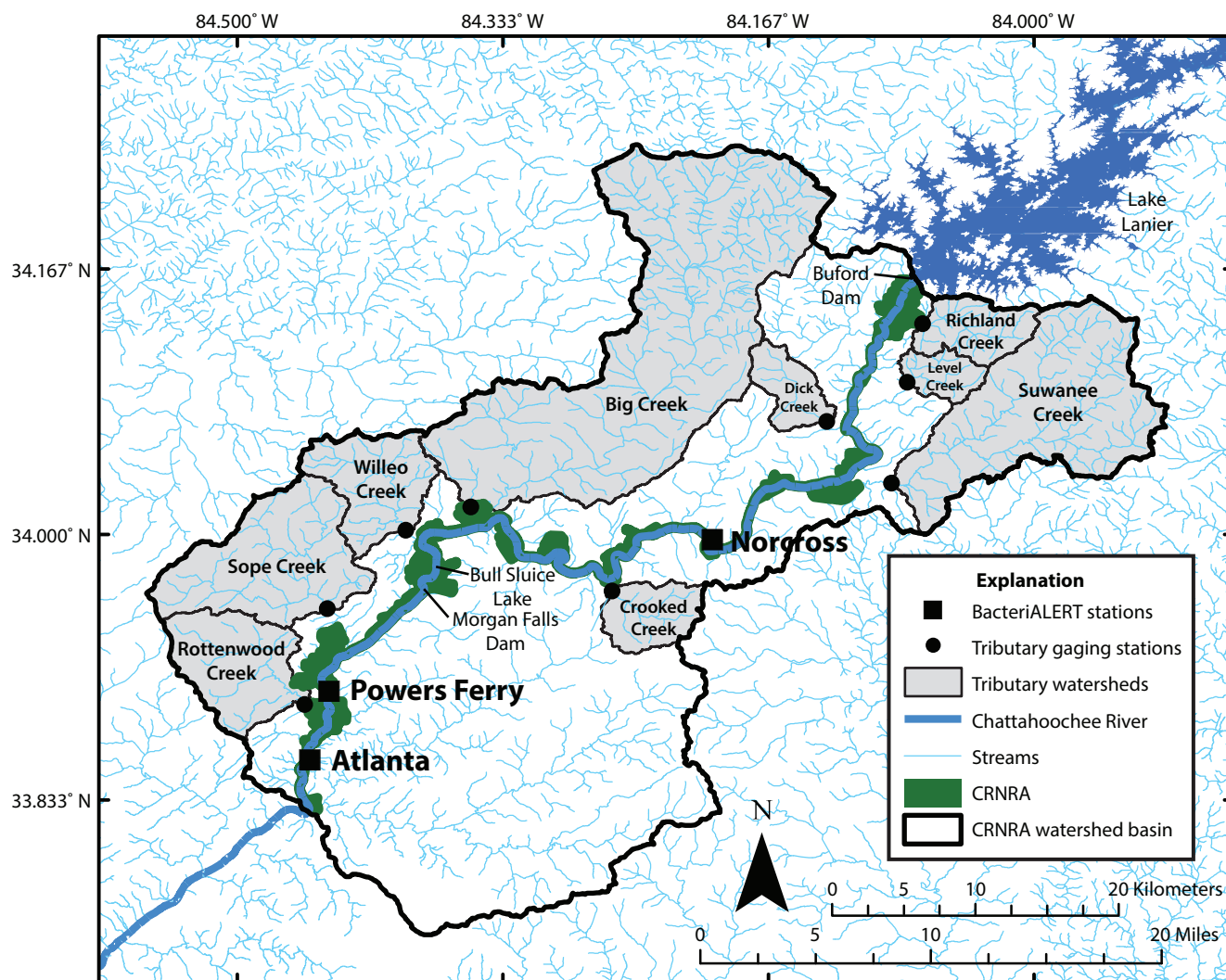


Figure 3. Map showing the Chattahoochee River National Recreation Area (CRNRA) and watershed basin, BacteriALERT program monitoring stations, and nine tributary watersheds monitored for streamflow.

adjoining gaged watersheds. To account for the travel time for tributary flows to reach the BacteriALERT stations, the sum of upstream tributary streamflows were calculated from prior 15-minute instantaneous streamflows using time-lags before the sample collected as designated in table 4. These time lags were calculated from the distance in river miles along the Chattahoochee River between the confluence of each tributary and BacteriALERT station of interest and a stream velocity of 3.8 feet/second (2.6 miles/hour) and resulted in similar lag times as reported in Lawrence (2012). The stream velocity was estimated based on the time it took for Buford Dam releases as observed from the station just below the dam to reach the stations at Norcross and Atlanta. However, there were some variations in stream velocities as velocities were slower when flows were lower. Timing to the Atlanta station was often confounded by dam operations at Morgan Falls. The portion of tributary flows upstream of Morgan Falls Dam to the

Powers Ferry and Atlanta stations are affected by the timing of operations of this dam and its source signal is also attenuated by the reservoir volume of Bull Sluice Lake, resulting in some non-ideal behavior in this model variable.

While tributary inflows can theoretically be estimated from the differences in flows from stations on the Chattahoochee River (with appropriate lag times for travel times), this was not done for various reasons. First, there is a tendency for differences to be inaccurate when the difference from tributary inflows is small relative to the larger flows of the Chattahoochee River used to calculate this difference. Second, the differences in flows between just below Buford Dam and the Powers Ferry and Atlanta stations may reflect more the timing of the Morgan Falls Dam operations and not the contribution of tributary inflows. The Morgan Falls Dam flow operations are not meant to mimic flow operations at Buford Dam despite the shallow nature and minimal storage

of Bull Sluice Lake; operations tend to reduce flow variations from releases at Buford Dam and the lake operates as a run-of-the-river reservoir only during the highest flows (Lawrence, 2012). Redundancy in streamflow variables was prevented by including only at-site and tributary flows in the models, as the flow from Buford Dam (with an appropriate lag) was a function of the difference in the other two variables.

The Powers Ferry gage did not have a stage-discharge rating curve developed at the time of this analysis, so at-site streamflow was estimated as the streamflow from the Atlanta station minus the additional intervening tributary streamflow from the Rottenwood Creek watershed. Modeling was done in two ways: (1) using time-lags for the two streamflows (table 4) which resulted in bacteria density predictions to be delayed by one hour, and (2) ignoring the lags to allow for the predictions to be done in real time. Model error estimates were compared to evaluate whether the model without lags could be employed without substantial degradation in prediction precision.

E. coli densities were modeled in logarithmic space (base 10) to help better linearize the relation with the explanatory variables and to address the unequal variance in densities observed across its range in arithmetic space that can result in poor model performance. Various transformations of model explanatory variables were explored to linearize their relations with the logarithm of *E. coli* densities, as necessary for multiple linear regression. For flow and turbidity variables, transformations explored included squared, logarithmic and logarithmic-squared, square root, and inverse relations. Logarithmic and logarithmic-squared transformations were found to work best across all models and were selected for the final model fitting. Limiting the number of transformations used in the model selection process is desirable as the transformed model parameters for a particular variable will be correlated with each other, which can confound the model selection process. The model relation with season was linearized by using a pair of sine and cosine functions of day of year, which allows for the model parameters to be fit using multiple linear regression while being mathematically equivalent to fitting a single sinusoidal function with a phase shift that requires a non-linear fit. The seasonal sine and cosine model terms worked as a single variable and as such were both included in a model as long as one term was deemed significant in the model fit. A linear trend term was fit with decimal year and represents a percentage change per unit time due to its fit in logarithmic space. The trend as a percentage change per year was calculated by determining any predicted value one-year in the future and calculating the percentage change between the predicted densities in arithmetic space. The initial, full equation used to fit regression model is

$$\begin{aligned} \log_{10}(Ecoli) = & a_0 + a_1 \cdot \log_{10}(Turb) + a_2 \cdot (\log_{10}(Turb))^2 + \\ & a_3 \cdot WTemp + a_4 \cdot (WTemp)^2 + a_5 \cdot \log_{10}(Q_{Stat}) + \\ & a_6 \cdot (\log_{10}(Q_{Stat}))^2 + a_7 \cdot \log_{10}(Q_{Trib}) + a_8 \cdot (\log_{10}(Q_{Trib}))^2 + \\ & a_9 \cdot \sin DOY + a_{10} \cdot \cos DOY + a_{11} \cdot DecYear \end{aligned} \quad (1)$$

where

$a_1..a_{11}$	coefficients of fitted parameters;
<i>Ecoli</i>	<i>E. coli</i> density, in colony forming units per 100 mL;
<i>Turb</i>	turbidity, in Formazin Nephelometric Units;
<i>WTemp</i>	water temperature, in degrees Celsius;
Q_{Stat}	at station streamflow, in cubic feet per second;
Q_{Trib}	upstream tributary streamflow, in cubic feet per second;
<i>DOY</i>	day of year; and
<i>DecYear</i>	decimal year.

Models were calibrated from samples collected between October 4, 2007, and May 23, 2019. The start date corresponds to when all nine tributaries consistently had flows available to use in the modeling. Model parameters were fit using ordinary least squares (OLS) approach in which the sum of squares in the differences between observed and predicated densities was minimized. A forward stepwise regression was used to determine which parameters were significant in equation 1 and used in the final models. The Bayesian Information Criteria (BIC; Schwarz, 1978) was used to determine the optimal set of model parameters. The BIC weighs the benefit of including additional parameters on model improvement versus the risk of overparameterization (overfitting). While overparameterization does not necessarily bias model parameter estimates, it can inflate estimated model variances thereby making model predictions appear less precise. The BIC was selected as opposed to the also commonly used Akaike Information Criteria (AIC; Akaike, 1974) because the BIC selects the model parameters that best explains the inherent variability in the data while the AIC selects the model parameters that best predicts the given calibration dataset. The BIC is therefore a more appropriate criteria because the developed models will be used in a future predictive manner as opposed to just predicting its calibration dataset. Of the two criteria, the BIC has a tendency to select fewer model parameters, though both criteria often result in selection of the same set of model parameters. Logarithm worth is an indication of the relative importance of various model parameters within a regression (larger values indicate greater importance) and was calculated from the t-ratio probabilities for each model parameter.

Models were fit and model statistics were calculated using JMP statistical data analysis software version 14.3.0 (JMP, 1989–2019).

Models developed in logarithmic space typically underestimate density because the model fits the geometric mean while the desired arithmetic mean is higher (Ferguson, 1986). A retransformation bias correction factor was calculated from the nonparametric Duan's smearing estimator (Duan, 1983), which is calculated as the mean of the model residuals (ϵ) transformed into arithmetic space:

$$\text{Duan's smearing estimator} = \frac{\sum_{i=1}^n 10\epsilon_i}{n} \quad (2)$$

where n is the number of observations (sample densities) and i represents the i th number of the sample residual. Model predicted densities were corrected by multiplying the predicted value transformed back into arithmetic space by the Duan's smearing estimator.

Outlier Analysis and Model Assessment

An outlier analysis of sample densities used in calibrating the regression models was performed to identify and remove values that may affect regression model fits. In the process of OLS regression, where the sum of the squares of the differences between observed and predicted densities are minimized, outliers can have a considerable influence on the model fit and result in biased model predictions and diminished model precision. Outliers were identified from plots of model residuals densities (observed minus model predicted densities) versus the predicted densities. Outliers were identified from the residuals because identification of outliers from plots of densities versus explanatory variables was often misleading due to the influence of multiple variable relations with *E. coli* densities. Outliers were identified and removed when their residual densities exceeded minimum and maximum cutoff values determined for each station. These cutoffs were set based on where residual densities occurred well outside of the margin of the cloud of points on the residual plot.

Models fits were evaluated from (1) the model coefficient of determination (model R^2), (2) model root-mean-square error (RMSE), (3) plots of observed versus predicted densities, and (4) checks of OLS regression assumptions through an analysis of the model errors (residuals). The model R^2 indicates the proportion of variance in *E. coli* densities that was explained by the model. The RMSE quantifies the precision of the model predictions. Observed versus predicted density plots illustrate how precisely and accurately the predicted values fit observed densities.

OLS regression provides unbiased estimates when errors are homoscedastic (have uniform variance) and independent. Furthermore, OLS provides the most probable/efficient model parameters when errors are distributed normally because the likelihood function is maximized. Therefore, model residual densities were assessed for (1) normality, (2) independence

(lack of serial correlation), and (3) to evaluate that they were homoscedastic when plotted versus the predicted densities and the explanatory variables used in the models. Normality of the residuals were assessed through (1) histograms which were compared to the normal distributions fitted to their distributions, (2) normal quantile (quantile-quantile) plots that show how closely the residuals follow a normal distribution, and (3) the Shapiro-Wilk W Test, a statistical test that determines if the residuals come from the normal distribution. A non-normal error distribution may bias the model parameter fit while the RMSE is calculated based on a normal distribution of errors. However, McCulloch and Neuhaus (2011) indicated that OLS parameter estimates were still robust for non-normal distributions that exhibited skew (unsymmetrical) or kurtosis (heavy-tailed) but were still unimodal. Serial correlation in errors were tested by calculating the autocorrelation in the residuals with a one-sample lag. Serial correlation coefficients can vary from -1 to 1 with values near zero indicating little correlation between residuals while values near +/-1 indicate high correlation. The presence of a high levels of serial correlation has several effects: (1) the model parameter coefficients are no longer the most efficient estimates possible (but are still unbiased; Helsel and Hirsch, 1992), (2) the significance of the model coefficients are overestimated, affecting the BIC criteria such that the optimal model might not be selected, and (3) uncertainty is underestimated resulting in confidence intervals that are too narrow. Model residuals were assessed by visually evaluating plots of residual densities versus the predicted densities and the model explanatory variables year, day of year, water temperature, at-site and upstream tributary streamflows, and turbidity. Residuals were assessed for homoscedasticity, such that their variance is about equal throughout the range of each variable. Unequal variance indicates that the RMSE is not constant, but dependent upon the magnitude of the predicted or explanatory variable. Residuals also were assessed to ensure that they vary equally above and below zero throughout the range of each variable. If residuals occur more frequently above or below zero over a portion of the range of predicted or explanatory variables, it is an indication of a poor model fit and that model predictions are biased under these conditions. A poor model fit could be the result of using inadequate explanatory variables or due to the use of the wrong parameter transformation to linearize the relation with *E. coli* density.

The effectiveness of the models to predict compliance and exceedance of the BAV was evaluated by comparing the model predicted *E. coli* densities with the observed sample densities over the model calibration period. This was evaluated by quantifying the percentage of occurrence of four cases: two consistent cases when predicted and observed densities either (1) both complied with or (2) both exceeded the BAV; along with two inconsistent cases when (3) the observed densities complied with the BAV while the predicted densities were not compliant (false positive), and (4) the observed densities exceeded the BAV while the predicted densities were compliant (false negative).

Results

Escherichia coli (*E. coli*) Monitoring

E. coli densities varied by at least three orders of magnitude at each of the three BacteriALERT stations during the monitoring period (fig. 4). The geometric mean of the *E. coli* densities was lowest at the Norcross station and highest at the Powers Ferry station (table 5). Plots comparing *E. coli* densities between the stations from samples collected on the same day show that while densities tend to be similarly high or low between stations, there is considerable variability (fig. 5). The Powers Ferry and Atlanta stations, which are the stations closest to each other (3.3 river miles apart; table 4) have more similar *E. coli* densities than with the Norcross station. Comparisons of observed *E. coli* densities between stations indicated that densities were sometimes higher and sometimes lower than each other across the range of observed densities; with the exception of Powers Ferry, in which *E. coli* densities were always higher than at Norcross when Norcross densities were above 1,000 cfu/100 mL.

Over the period of study, the EPA Beach Action Value (BAV) of 235 cfu/100 mL was exceeded in 15.5, 33.6, and 30.3 percent of the samples collected at the Norcross, Powers Ferry, and Atlanta stations, respectively (table 6; fig. 4). While samples collected from the BacteriALERT program are not used to determine whether the Recreational Water Quality Criteria (RWQC) was met, and while the criteria is based the exceedance of a minimum of four samples collected over a 30-day period (see sidebar “Georgia Water Quality Standards for Recreation and Drinking Water Use Designations” for details), it is instructive to compare the frequency that individual sample densities exceed the geometric mean (GM) and

statistical threshold values (STV) used in this criteria (table 6). The frequency of exceedance of individual samples of the GM (23.9 to 46.1 percent) and STV (10.6 to 23.1 percent) criteria indicates a likelihood of at least occasional exceedance of the RWQC criteria at all three stations.

Plots of *E. coli* densities versus possible model explanatory variables illustrate some of their relations (figs. 6 to 8). Data are shown only for sample densities used for calibrating the regression models, for the period October 4, 2007, to May 23, 2019. These plots can sometimes be difficult to interpret due to the case that some explanatory variables are correlated with each other and they show relations with individual variables while the relations are due to a combination of multiple variables. *E. coli* densities appear to be increasing over time at the Powers Ferry station (fig. 7A) while any trends at the other two stations are not visually apparent. The Powers Ferry station has a shorter record as sampling was initiated in 2016. The trend in increasing densities at Powers Ferry could be the result of its relation of higher densities with higher streamflows combined with increasing wet climate conditions during this period. All three stations exhibited seasonal patterns with densities being highest in the summer, peaking around day of year 200 to 220 (July 20th to August 9th). The seasonal patterns were most distinct at the Norcross station (fig. 6B).

E. coli densities increased with increasing water temperature at Norcross and Atlanta stations (figs. 6C and 8C, respectively), while no obvious relation was evident at the Powers Ferry Station. *E. coli* densities generally increased with increasing streamflow, particularly for increases in upstream tributary streamflows. Some of the observed density relations with at-site streamflow is the result of contributions of flow from upstream tributaries. This is particularly apparent for the

Table 5. BacteriALERT *Escherichia coli* (*E. coli*) sampling through May 23, 2019, including dataset start date and statistical distribution of densities.

[Datasets: Monitoring, entire sample dataset; Calibration, model calibration dataset for developing regression models. Abbreviations: cfu, colony forming units; Ga., Georgia; mL, milliliters]

USGS station number	Station name	Dataset	Dataset start date	Number of samples (n)	Minimum (cfu/100 mL)	Geometric mean (cfu/100 mL)	90th percentile (cfu/100 mL)	Maximum (cfu/100 mL)
02335000	Chattahoochee River near Norcross, Ga.	Monitoring	Oct. 23, 2000	1,969	<1	293	452	18,000
		Calibration	Oct. 4, 2007	518	6	211	300	8,700
02335880	Chattahoochee River at Powers Ferry and I-285 near Atlanta, Ga.	Monitoring	May 5, 2016	134	13	592	1,600	13,000
		Calibration	June 16, 2016	97	17	669	1,620	13,000
02336000	Chattahoochee River at Atlanta, Ga.	Monitoring	Oct. 23, 2000	1,938	7	529	1,100	28,000
		Calibration	Oct. 4, 2007	380	12	381	818	9,200

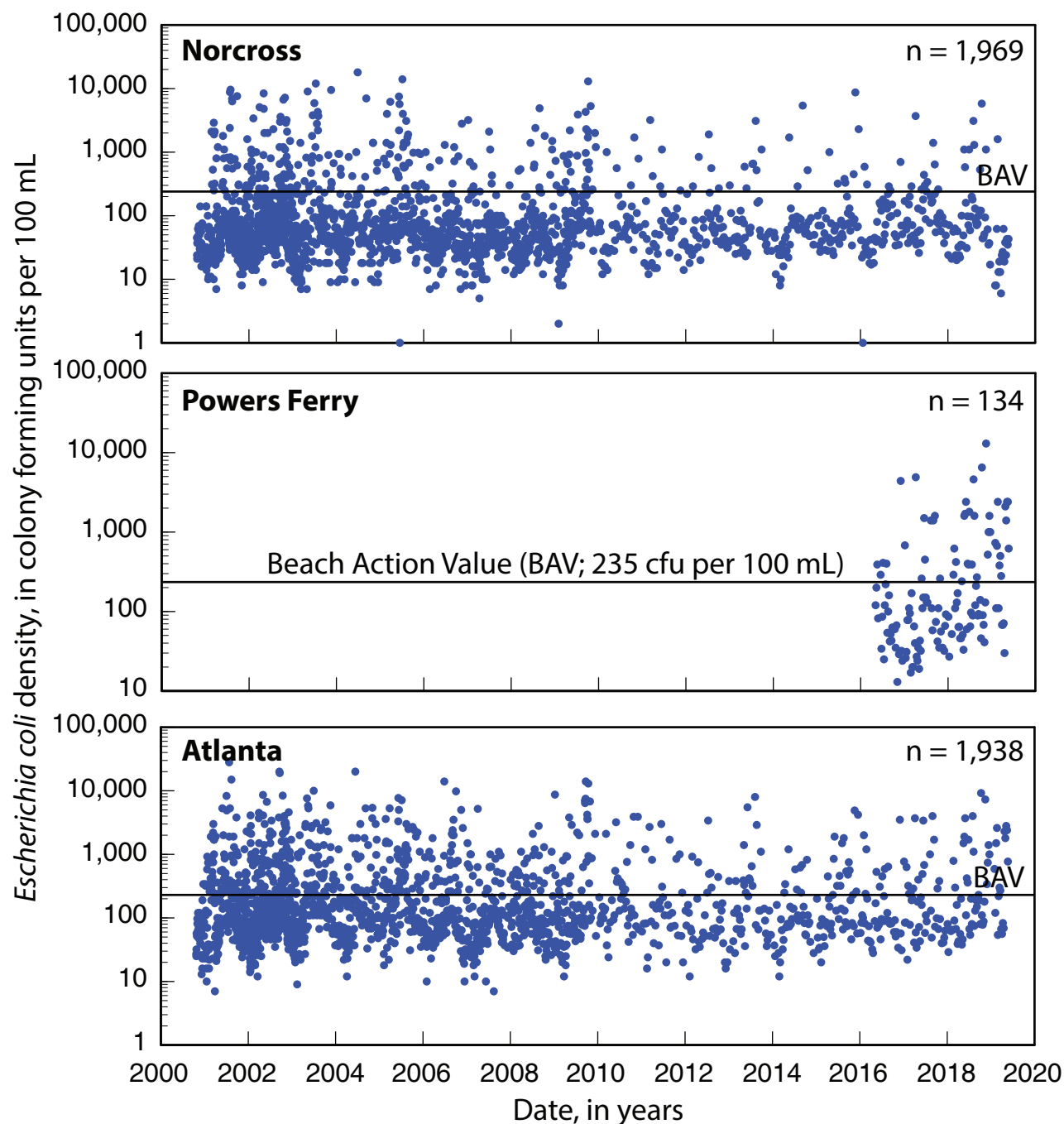


Figure 4. Sample *Escherichia coli* (*E. coli*) densities on the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 23, 2000, to May 23, 2019 (May 5, 2016, to May 23, 2019, for Powers Ferry). Abbreviations: cfu per 100 mL, colony forming units per 100 milliliters; n, number of samples (each plotted as a blue dot).

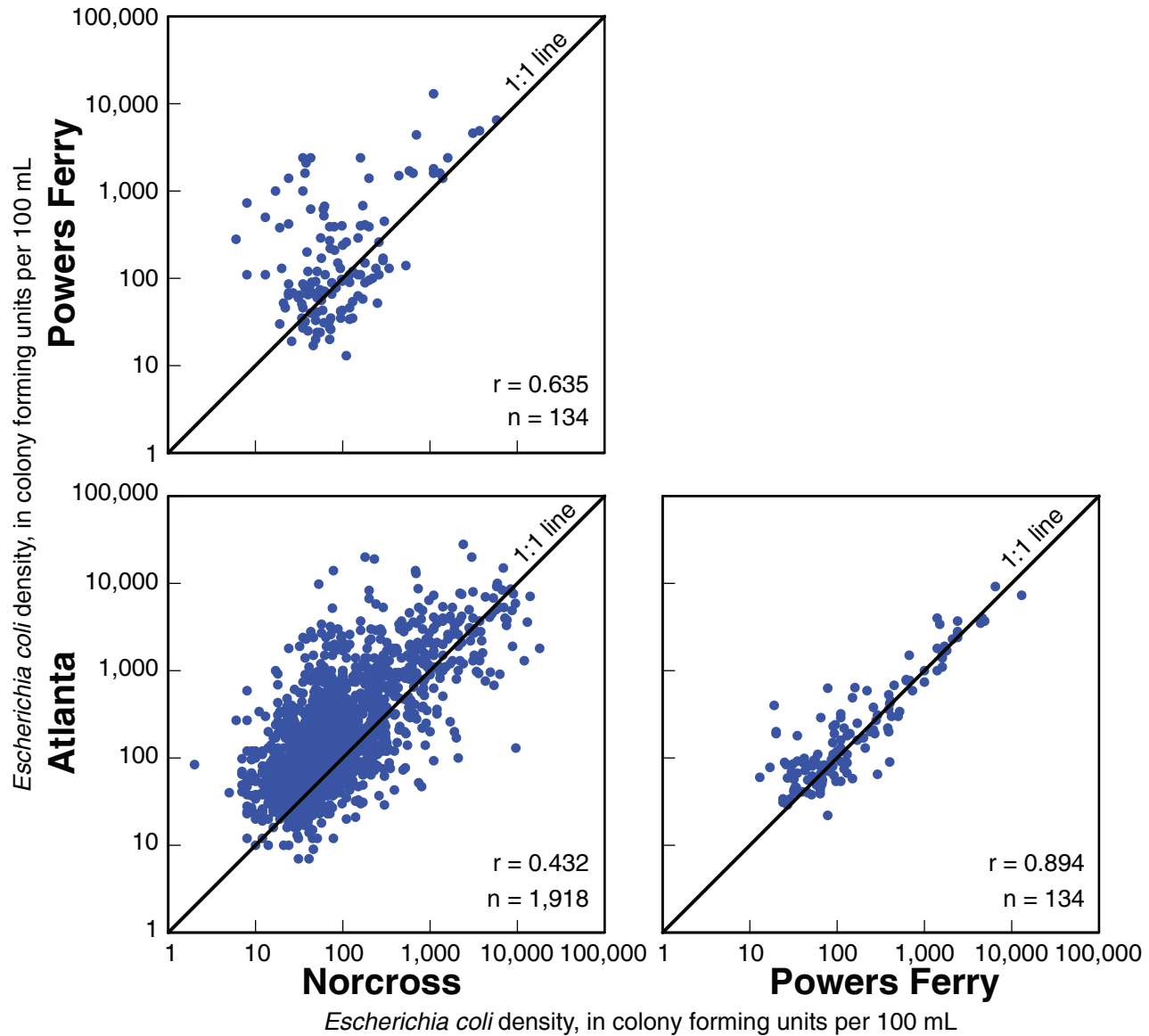


Figure 5. Scatterplots of *Escherichia coli* (*E. coli*) densities from samples collected on the same day between stations on the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 23, 2000, to May 23, 2019 (May 5, 2016, to May 23, 2019, for Powers Ferry). Abbreviations: mL, milliliters; n, number of samples (blue dots); r, Pearson product-moment correlation coefficient.

Norcross station at higher streamflows, in which higher *E. coli* densities are observed when streamflow is predominantly from upstream tributaries while lower *E. coli* densities are observed when streamflow is predominantly from water released from Buford Dam (fig. 6D). Strong relations of increasing *E. coli* densities with increasing turbidity were observed for all three stations (figs. 6E, 7E, and 8E).

The range of values of the model explanatory variables within the sample calibration dataset should span the range in conditions observed for the explanatory variables in order to avoid extrapolation of regression model predictions that can result in additional uncertainty. Comparisons of distributions

of calibration sample explanatory variable values with the distributions of observed 15-minute time-series values indicate that the calibration datasets sufficiently cover the ranges in observed turbidity, at-site streamflow, and upstream tributary streamflow (fig. 9). High outliers indicated in the calibration samples box plots are not actually outliers but represent the long tail of rarely occurring values during high-flow conditions. Other explanatory variables not shown in figure 9, decimal-year, day-of-year, and water temperature, should be sufficiently represented due to the fixed-interval sampling approach used in this study.

Table 6. Number of samples exceeding the Environmental Protection Agency's (EPAs) Beach Action Value (BAV) of 235 cfu/100 mL *E. coli*, the geometric mean (GM) of 126 cfu/100 mL *E. coli*, and the statistical threshold value (STV) of 410 cfu/100 mL *E. coli* for the period October 23, 2000, to May 23, 2019 (May 5, 2016, to May 23, 2019, for Powers Ferry).

[GM and STV criteria are from the EPA recommended 2012 Recreational Water Quality Criteria (RWQC; [table 1](#)). Abbreviations: cfu/100 mL, colony forming units per 100 milliliters; Ga., Georgia]

Station name	Number of samples (n)	Number of samples exceeding the BAV	BAV exceedance (percent)	Number of samples exceeding the GM	GM exceedance (percent)	Number of samples exceeding the STV	STV exceedance (percent)
Chattahoochee River near Norcross, Ga.	1,969	305	15.5	471	23.9	208	10.6
Chattahoochee River at Powers Ferry and I-285 near Atlanta, Ga.	134	45	33.6	58	43.3	31	23.1
Chattahoochee River at Atlanta, Ga.	1,938	587	30.3	893	46.1	404	20.8

Outlier Analysis

A summary of the criteria and the number of sample densities identified as outliers and removed from the model calibration dataset are summarized in [table 7](#). The outlier criteria, as defined by the exceedance of residual *E. coli* densities thresholds, ranged from ± 0.6 to ± 0.8 log-units. Of the 1,043 samples considered for calibration of the regression models, 40 samples (3.8 percent) were identified as outliers with 82.8 percent of these identified as high outliers. The Powers Ferry station had the highest percentage of outliers at 7.6 percent. The prevalence of high outliers resulted in predominantly false negative misidentification of compliance of the BAV (62.5 percent of the outlier samples), in which predicted densities were below the BAV whereas the sample outlier density exceeded the BAV ([table 1.1](#), appendix 1). False positives occurred for 2.5 percent of the sample outliers, while both predicted and outlier densities complied with the BAV in 15.0 percent of the samples and exceeded the BAV in 20.0 percent of the samples.

Some spatial and temporal patterns in outlier occurrence were observed. Five of eight sample outliers at Powers Ferry were also high outliers at the Atlanta station for those sample dates, while outliers did not occur on the same sample dates at the Norcross and Atlanta stations. ([table 1.1](#), appendix 1). High outliers occurred at both Powers Ferry and Atlanta stations for the four consecutive weekly samples collected at each station over the period April 25, 2019, to May 16, 2019, suggesting that there is likely a single cause for elevated *E. coli* densities above model predicted densities that lasted over an extended period of time and affected both stations.

Escherichia coli (*E. coli*) Models

A total of five *E. coli* regression models were developed for the three stations ([table 8](#)). Two models were developed for the Norcross station; one included a time trend parameter (decimal year) whereas a second model did not. Two models were developed for the Powers Ferry station; one model included the use of time lags for the calculation of at-site streamflow (flows at Atlanta minus intervening Rottenwood Creek tributary inflow) whereas a second model ignored these lags. The Powers Ferry station models did not include time trend parameters due to its short monitoring period of three years. Only one model was fit for the Atlanta station because there was not a significant time trend model parameter. Models were calibrated using 518 samples at Norcross, 97 samples at Powers Ferry, and 380 samples at Atlanta ([table 8](#)), which should be a sufficient number of samples for modeling purposes.

The forward stepwise regression using the BIC criteria resulted in the selection of six to eight model parameters for each model out of the possible 12 parameters included in [equation 1](#) (including the model intercept; [table 9](#)). The model parameters used in each model and their coefficients are included in [table 9](#). All models have log-turbidity and (or) squared log-turbidity parameters with positive model coefficients indicating a positive relation between *E. coli* density and turbidity. This was expected due to the association of a substantial portion of bacteria with sediment (Fries and others, 2006; Krometis and others, 2007) combined with the positive relation between sediment concentration and turbidity. Norcross and Powers Ferry models include at-site streamflow

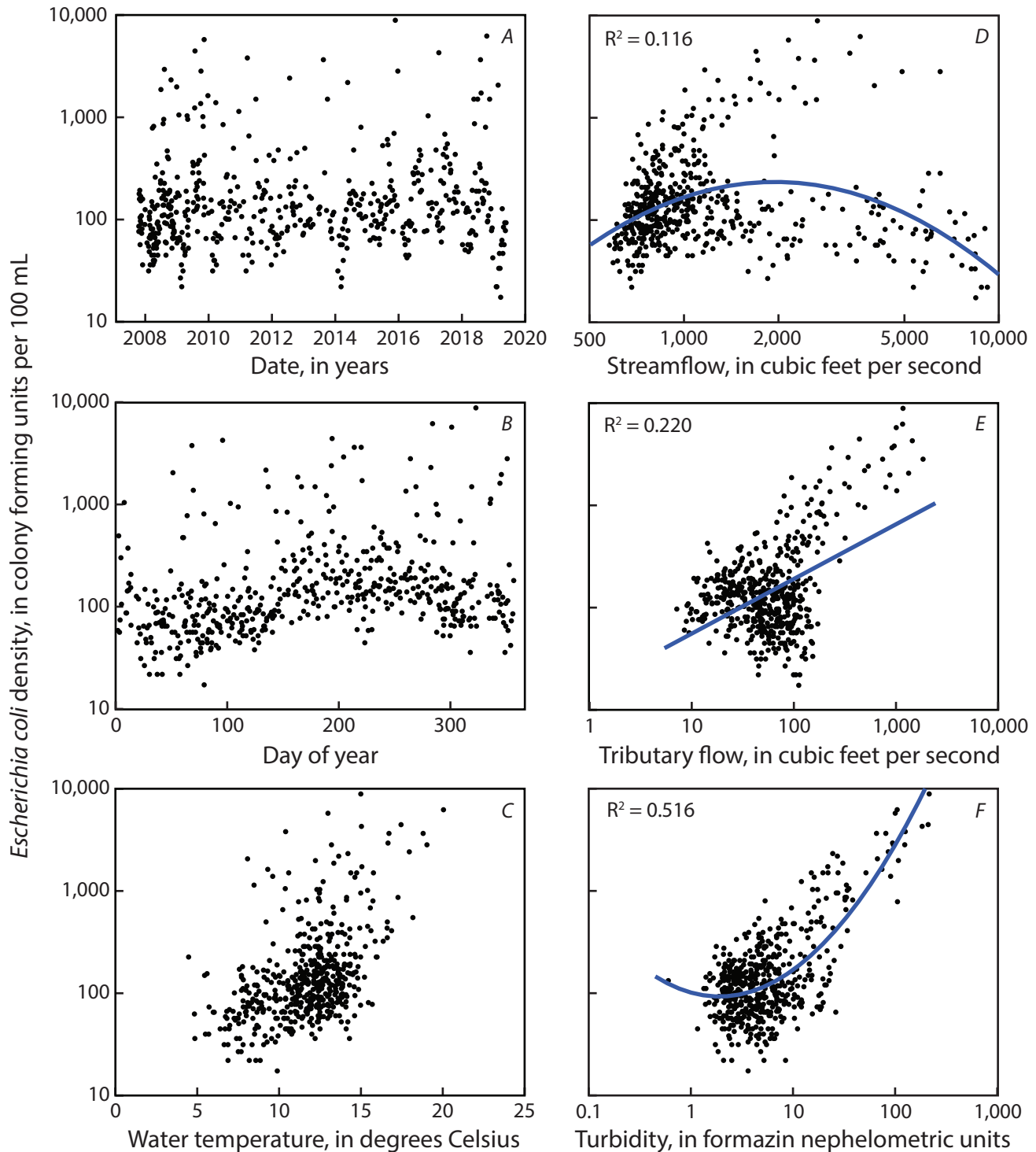


Figure 6. Graphs showing *Escherichia coli* (*E. coli*) density versus A, time; B, season; C, water temperature; D, streamflow; E, upstream tributary streamflow; and F, turbidity at Chattahoochee River near Norcross, Ga. (USGS station 02335000) for the period October 1, 2007, to May 23, 2019. Linear or quadratic relations (blue line) and relation R-squared (R^2) values are shown for parameters used to model *E. coli*. Black dots are samples ($n=1,969$). Abbreviations: mL, milliliters.

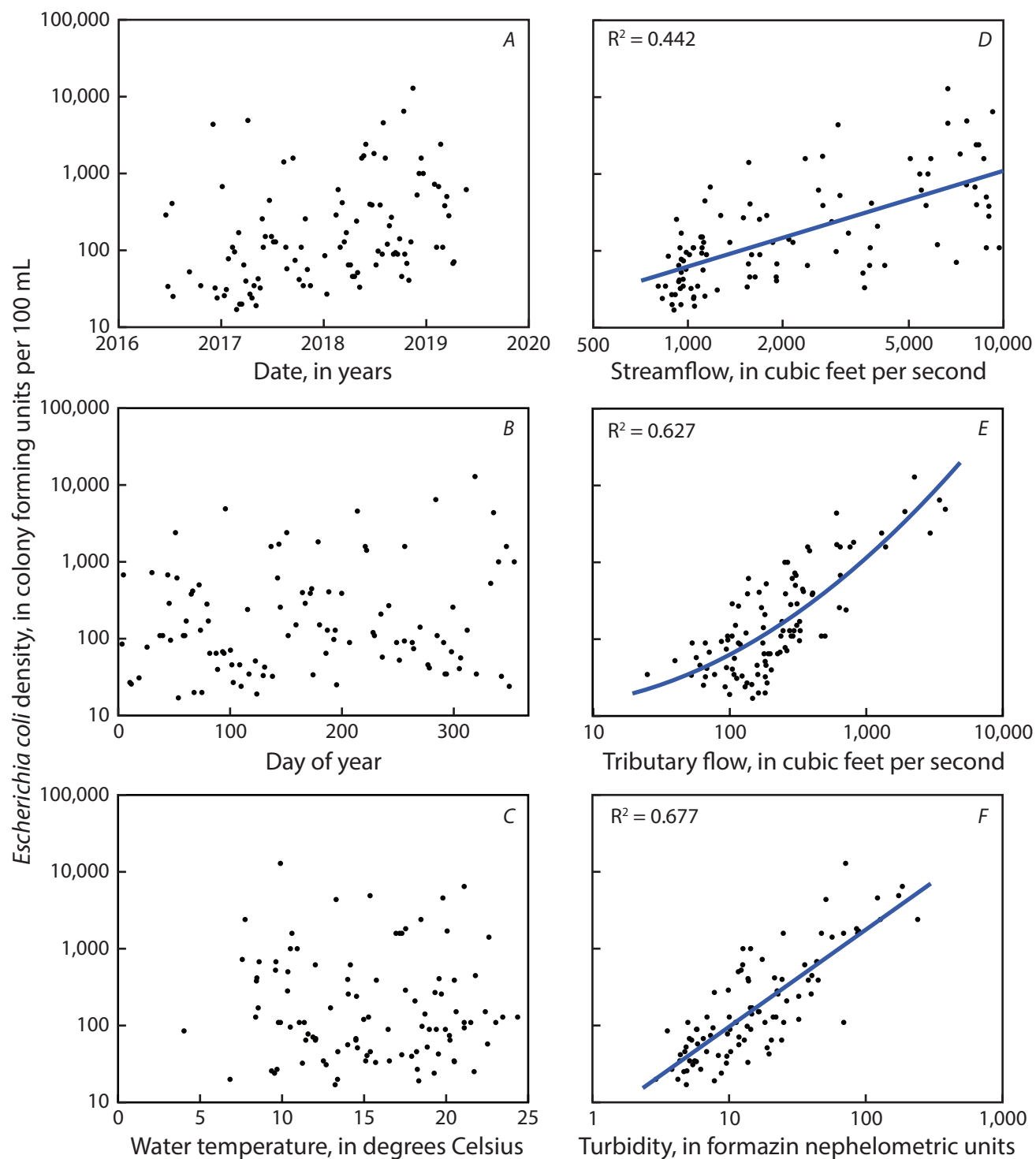


Figure 7. Graphs showing *Escherichia coli* (*E. coli*) density versus A, time; B, season; C, water temperature; D, streamflow; E, upstream tributary streamflow; and F, turbidity at Chattahoochee River at Powers Ferry and I-285 near Atlanta, Ga. (USGS station 02335880) for period June 16, 2016, to May 23, 2019. Linear or quadratic relations (blue line) and relation R-squared (R^2) values are shown for parameters used to model *E. coli*. Black dots are samples ($n=134$). Abbreviations: mL, milliliters.

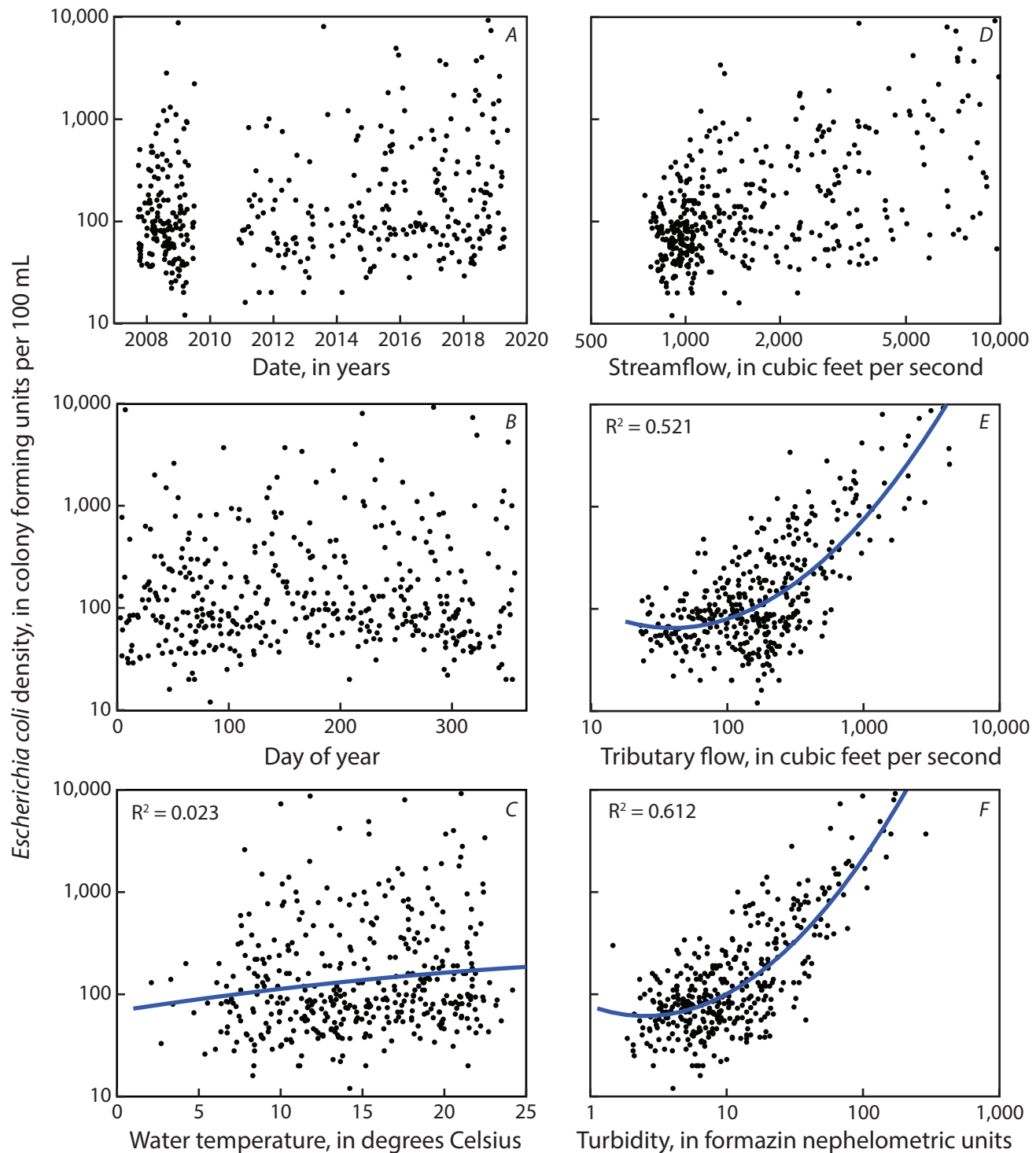


Figure 8. Scatterplots showing *Escherichia coli* (*E. coli*) density versus A, time; B, season; C, water temperature; D, streamflow; E, upstream tributary streamflow; and F, turbidity at Chattahoochee River at Atlanta, Ga. (USGS station 02336000) for the period October 4, 2007, to May 23, 2019. Linear or quadratic relations (blue line) and relation R-squared (R^2) values are shown for parameters used to model *E. coli*. Black dots are samples ($n=1,938$). Abbreviations: mL, milliliters.

Figure 9. Comparison plots of 15-minute interval time-series versus calibration sample distributions of the model explanatory variables (turbidity, at-site streamflow, and upstream tributary streamflow) for the stations on the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 1, 2007, to May 30, 2019. Distribution probabilities are for 0.1 interval classes. Abbreviations: IQR, interquartile range (between 25th and 75th percentiles); n, the number of 15-minute intervals of time-series or number of samples; %, percent.

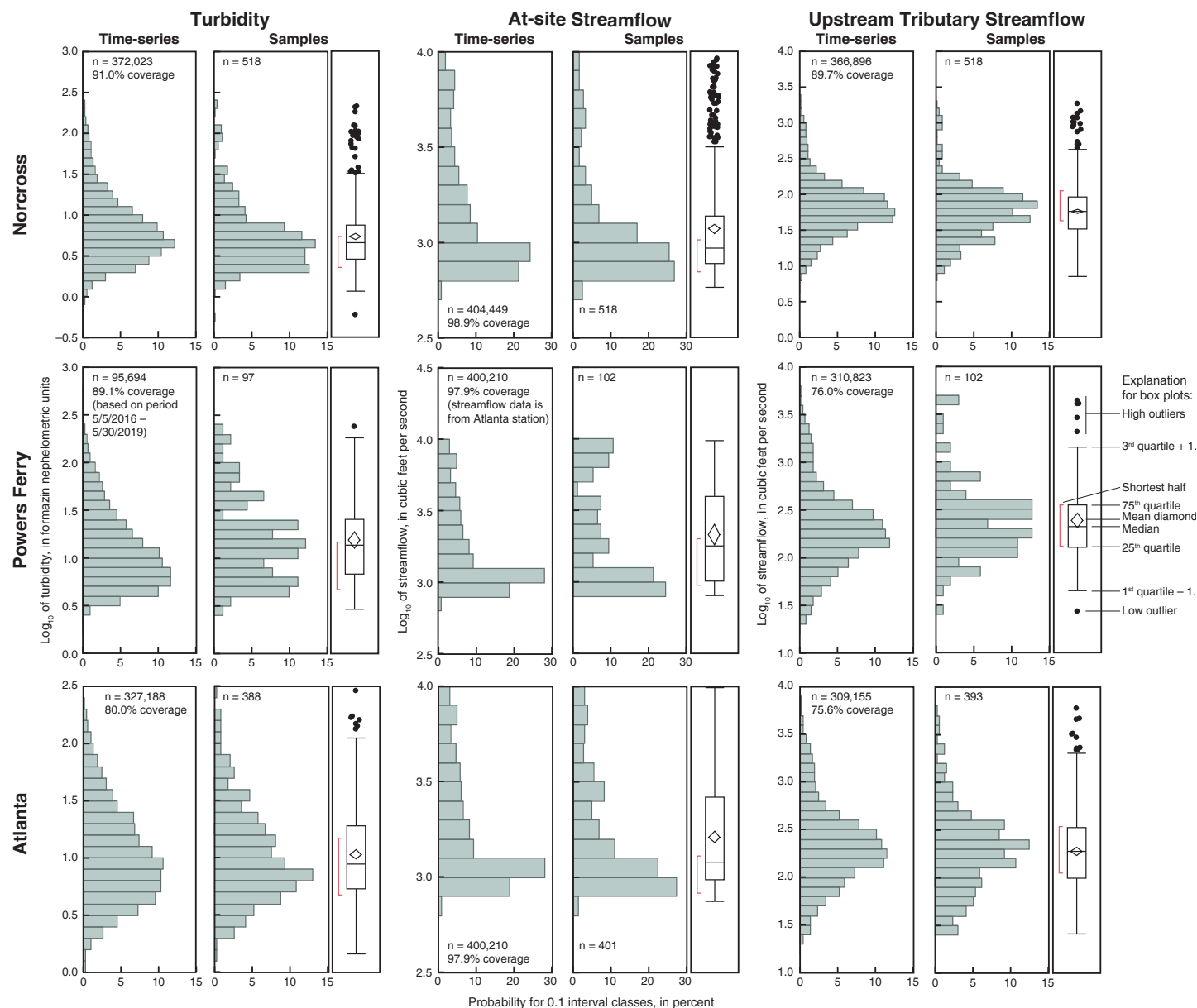


Table 7. Summary of outliers removed from regression analysis, for the period October 4, 2007, to May 23, 2019 (June 16, 2016, to May 23, 2019, for Powers Ferry).

[Abbreviations: cfu, colony forming units; mL, milliliters]

Station	Residual criteria in $\log_{10}(E. coli)$ in cfu per 100 mL)	Total number of samples	Number of outliers removed	Percentage of outliers (percent)	Number of high outliers	Number of low outliers
Norcross	$<-0.6; >0.6$	536	18	3.4	15	3
Powers Ferry	$<-0.7; >0.7$	105	8	7.6	6	2
Atlanta	$<-0.75; >0.8$	401	14	3.5	12	2
Total		1,042	40	3.8	33	7

Table 8. Summary of *Escherichia coli* (*E. coli*) model statistics and the Duan's smearing estimator.[Residual normality test using Shapiro-Wilk W Test; *p*-values <0.05 indicate that data are not from a normal distribution]

Station	Model	Number of samples	Number of model parameters	Degrees of freedom (DFE)	Model adjusted R-squared	Root-mean-square error (RMSE)	Auto correlation of residuals	Residual normality test probability (<i>p</i> -value)	Duan's smearing estimator
Norcross	No time trend	518	7	511	0.791	0.231	0.135	0.0029	1.157
	With time trend	518	8	510	0.795	0.229	0.105	0.0019	1.154
Powers Ferry	Downstream flows lagged	97	6	91	0.783	0.310	0.011	0.3235	1.245
	Downstream flows not lagged	97	6	91	0.784	0.309	0.005	0.4490	1.245
Atlanta	No time trend	380	6	374	0.686	0.292	0.279	0.0449	1.240

Table 9. *Escherichia coli* (*E. coli*) model parameter coefficients.

[Abbreviations: °C, degrees Celsius; cfs, cubic feet per second; FNU, Formazin Nephelometric Units; Log, logarithm base 10; — parameter not significant; na, not applicable]

Station	Model	Intercept	Log turbidity (FNU)	Log turbidity squared (FNU)	Water temperature squared (°C)	Log at-site stream-flow (cfs)	Log at-site stream-flow squared (cfs)	Log up-stream tributary stream-flow (cfs)	Log up-stream tributary stream-flow squared (cfs)	Sine day-of-year	Cosine day-of-year	Trend (decimal year)
Norcross	No time trend	-7.542	—	0.3265	—	5.444	-0.8879	0.4958	—	-0.2433	-0.2395	na
	With time trend	-26.35	—	0.3450	—	5.102	-0.8416	0.4537	—	-0.2363	-0.2309	0.009680
Powers Ferry	Downstream flows lagged	-0.6582	0.5319	—	—	0.4459	—	—	0.1315	-0.2343	0.0032	na
	Downstream flows not lagged	-0.6540	0.5374	—	—	0.4414	—	—	0.1320	-0.2346	0.0087	na
Atlanta	No time trend	1.115	—	0.1959	0.0003123	—	—	—	0.1308	-0.1635	-0.0439	na

Table 10. *Escherichia coli* (*E. coli*) model parameter estimates and statistics.[Logarithm worth is calculated from parameter t-ratio probability as $-\log_{10}(p\text{-value})$. Abbreviations: Log, logarithm base 10; na, not applicable]

Station	Model	Parameter	Estimate	Standard error	t ratio	Probability > t	Logarithm worth
Norcross	No time trend	Intercept	-7.541807	1.366627	-5.52	<.0001	na
		Log turbidity squared	0.326531	0.018712	17.45	<.0001	53.111
		Sine day-of-year	-0.243330	0.016585	-14.67	<.0001	40.198
		Cosine day-of-year	-0.239486	0.015912	-15.05	<.0001	41.914
		Log at-site streamflow	5.443907	0.843231	6.46	<.0001	9.602
		Log at-site streamflow squared	-0.887911	0.127628	-6.96	<.0001	10.971
		Log upstream tributary streamflow	0.495829	0.045812	10.82	<.0001	23.993
	With time trend	Intercept	-26.35231	6.051118	-4.35	<.0001	na
		Log turbidity squared	0.344965	0.019426	17.76	<.0001	54.558
		Sine day-of-year	-0.236288	0.016586	-14.25	<.0001	38.283
		Cosine day-of-year	-0.230857	0.016001	-14.43	<.0001	39.09
		Log at-site streamflow	5.101549	0.842627	6.05	<.0001	8.563
		Log at-site streamflow squared	-0.841588	0.127329	-6.61	<.0001	10.012
		Log upstream tributary streamflow	0.453690	0.047290	9.59	<.0001	19.423
Powers Ferry	Downstream flows lagged	Decimal year	0.009680	0.003035	3.19	0.0015	2.82
		Intercept	-0.658186	0.344689	-1.91	0.0593	na
		Log turbidity	0.531938	0.156601	3.40	0.0010	2.995
		Sine day-of-year	-0.234326	0.046670	-5.02	<.0001	5.595
		Cosine day-of-year	0.003225	0.053305	0.06	0.9519	0.021
		Log at-site streamflow	0.445919	0.123102	3.62	0.0005	3.319
		Log upstream tributary streamflow squared	0.131500	0.029036	4.53	<.0001	4.746
	Downstream flows not lagged	Intercept	-0.653950	0.336413	-1.94	0.055	na
		Log turbidity	0.537450	0.155134	3.46	0.0008	3.09
		Sine day-of-year	-0.234634	0.046510	-5.04	<.0001	5.637
		Cosine day-of-year	0.008725	0.052904	0.16	0.8694	0.061
		Log at-site streamflow not lagged	0.441389	0.119120	3.71	0.0004	3.441
		Log upstream tributary streamflow squared	0.132049	0.028943	4.56	<.0001	4.803
Atlanta	No time trend	Intercept	1.114912	0.076097	14.65	<.0001	na
		Log turbidity squared	0.195916	0.026830	7.30	<.0001	11.765
		Water temperature squared	0.000312	0.000260	1.20	0.2299	0.638
		Sine day-of-year	-0.163535	0.029740	-5.50	<.0001	7.149
		Cosine day-of-year	-0.043880	0.052919	-0.83	0.4075	0.39
		Log upstream tributary streamflow squared	0.130779	0.013800	9.48	<.0001	18.512

parameters while the Atlanta model did not include this variable. For the Norcross models, squared at-site streamflow parameters had negative coefficients, which indicated a complex relation where *E. coli* densities were lower at low and high streamflows and higher at intermediate streamflows (fig. 6D). All the models have log-streamflow and (or) squared log-streamflow parameters of tributary inflows with positive model coefficients, indicating that *E. coli* densities are higher when tributary flows are higher.

All models included seasonal terms, though the seasonal cosine parameter was not selected in the forward stepwise regression approach for the Powers Ferry and Atlanta models but were added to complete the seasonal function. The seasonal model coefficients indicated that *E. coli* densities were typically higher in late summer and lower in late winter. However, there were some differences in timing among the three stations. *E. coli* densities were highest about August 17th for the Norcross models, about October 2nd for the Powers Ferry models, and about September 16th for the Atlanta model. The amplitude of the seasonality calculated from the model coefficients indicated that the seasonal effect on *E. coli* densities was highest for Norcross and lowest for Atlanta. However, the Atlanta model also included a streamwater temperature parameter, which also varied seasonally. The Norcross model has a significant increasing trend with time. The coefficient of the trend parameter indicates an increase in density of 2.3 percent per year. This trend is substantial and necessitates using the model with the trend parameter for predicting *E. coli* densities accurately and to allow for this model to be operational in the future.

Model Assessment

Statistical analyses for each parameter estimate are summarized in table 10. Model parameters were not selected by the significance of t-tests of individual parameter but by the parameter selection criteria of minimizing the BIC of the set of model parameters. Logarithm worth indicated that, for the Norcross models, the most important variables (from

most to least) were turbidity, season, and upstream tributary streamflows. For the Powers Ferry models, the most important variables (from most to least) were season, upstream tributary streamflows, and at-site streamflows. For the Atlanta model, the most important variables (from most to least) were upstream tributary streamflows, turbidity, and season.

All *E. coli* density models explained a substantial amount of the variance, with model adjusted R²s ranging from 0.686 (Atlanta) to 0.795 (Norcross with time trend; table 8). The two Powers Ferry models had very similar model performances (as indicated by model adjusted R²s and RMSEs), indicating that the model in which downstream flows were not lagged is sufficient for real-time predictions, which avoids the 1-hour delayed predictions by using the model with the 1-hour forward lag-time associated with using the downstream Atlanta station streamflow for at-site streamflow. The model RMSEs range from 0.229 (Norcross with time trend) to 0.310 (Powers Ferry with downstream flows lagged). The RMSEs are difficult to assess due to their calculation in log-space. Confidence intervals in arithmetic-space were calculated based on RMSEs and assuming that the residuals are normally distributed and result in errors that are a percentage of the predicted value due to the mathematic retransformation (table 11). Confidence intervals are provided for two commonly reported standards, (1) ± 1 standard deviation and (2) the 95-percent confidence interval (calculated as ± 1.96 standard deviations based on a normal distribution). The ± 1 standard deviation standard indicates that values are expected to be within this range 68.27 percent of the time with occurrence in this range considered “merely probable,” while the 95-percent confidence interval is the expected range for 95 percent of the values with occurrence in this range considered “nearly certain.” While the error was distributed equally above and below the predicted values in log-space, the error in arithmetic space is wider above the predicted values. The lower and upper confidence intervals for the BAV of 235 cfu per 100 mL for ± 1 standard deviation are also reported for the various models. Using this standard, the range for the lower confidence limit is from 115 (both Powers Ferry models) to 139 (Norcross with time trends)

Table 11. *Escherichia coli* (*E. coli*) model confidence intervals calculated from the model root-mean-square error for ± 1 standard deviation (SD), 95-percent confidence intervals, and for the Beach Action Value (BAV) of 235 colony forming units (cfu) per 100 mL using ± 1 SD.

Station	Bacteria model	Root-mean-square error (RMSE)	-1 SD (percent)	+1 SD (percent)	Lower 95 percent confidence interval (percent)	Upper 95 percent confidence interval (percent)	BAV lower confidence interval (-1 SD; cfu/100 mL)	BAV upper confidence interval (+1 SD; cfu/100 mL)
Norcross	No time trend	0.231	-41.3	70.3	-65.5	190	138	400
	With time trend	0.229	-41.0	69.5	-65.2	187	139	398
Powers Ferry	Downstream flows lagged	0.310	-51.0	104	-76.0	316	115	479
	Downstream flows not lagged	0.309	-50.9	104	-75.9	314	115	478
Atlanta	No time trend	0.292	-49.0	96.1	-74.0	284	120	461

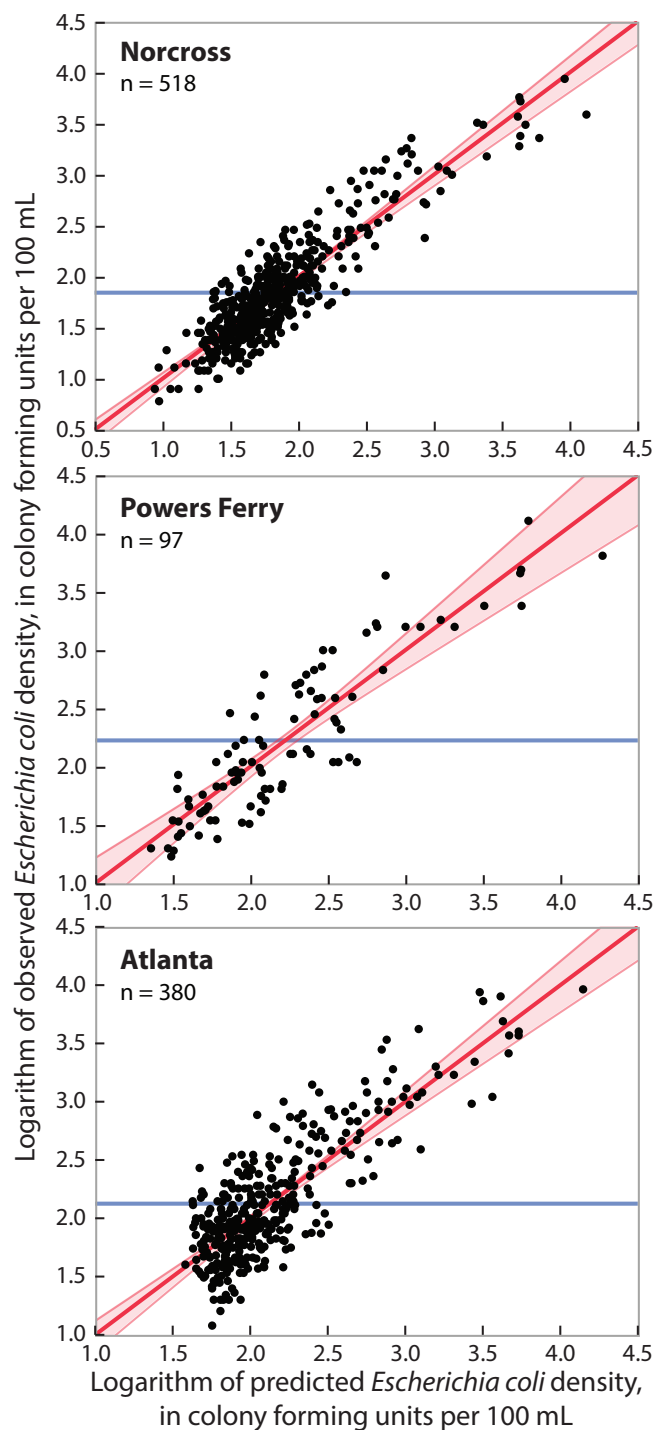


Figure 10. Scatterplots of observed versus model predicted *Escherichia coli* (*E. coli*) densities for the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 4, 2007, to May 23, 2019 (June 16, 2016, to May 23, 2019, for Powers Ferry). On each scatterplot, the thick red line is line of fit and the red shaded area indicates significance at p -value = 0.05; horizontal blue line represents mean density; black dots are samples used in model calibration. Logarithms are in base 10. Abbreviation: mL, milliliters; n, number of samples.

cfu per 100 mL, and the range for the upper confidence limit is from 398 (Norcross with time trends) to 479 (Powers Ferry) cfu per 100 mL. Due to uncertainty in the models, predicted *E. coli* densities at the BAV will be exceeded about 50 percent of the time; and predicted densities at one standard deviation below the BAV will still exceed the BAV about 15.9 percent of the time. The overall ranges are large and reflect how precisely the models can predict exceedance of the BAV criteria.

Duan's smearing estimator ranged from 1.154 to 1.245 and correspond to underestimates of *E. coli* densities from 13.3 to 19.7 percent, respectively, if these correction factors are not applied. These underestimates are substantial and are likely a result from the large degree of scatter about the regression (Ferguson, 1986).

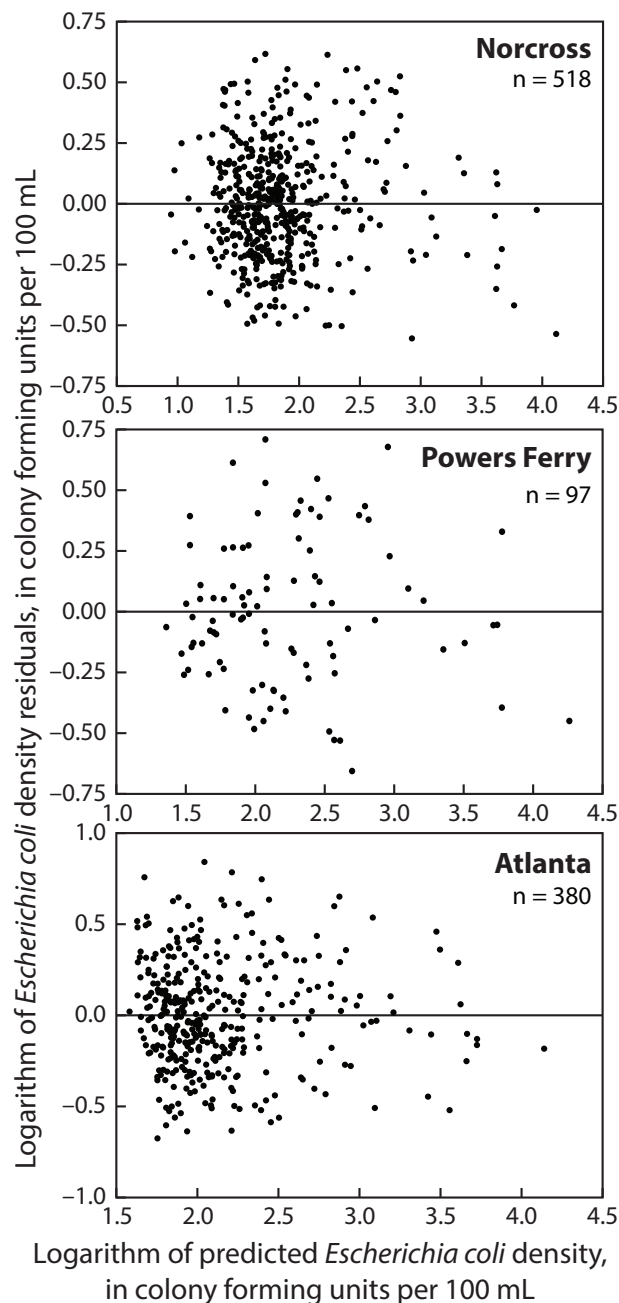


Figure 11. Scatterplots of *Escherichia coli* (*E. coli*) density model residuals versus model predicted densities for the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 4, 2007, to May 23, 2019 (June 16, 2016, to May 23, 2019, for Powers Ferry). Black dots are samples used in model calibrations. Logarithms are in base 10. Abbreviation: mL, milliliters; n, number of samples.

Bias in model predictions was assessed through the evaluation of plots of observed versus model predicted densities (fig. 10) and model residual densities versus model predicted values (fig. 11) and explanatory variables (figs. 12 to 14). Predicted and residual densities in these plots are from the Norcross and Atlanta models without time trends and the Powers Ferry model with downstream flows lagged. Plots for the other Norcross and Powers Ferry models were not discernably different than these stations' other models and are not shown. Observed versus model predicted plots show that the data fit closely to the one-to-one line (fig. 10). Plots of model residual densities versus predicted values indicate that residuals are distributed fairly equally above and below

zero over the full range of predictions (fig. 11). These observations indicate that bias in the model predictions is low. There appears to be a tendency to slightly overpredict *E. coli* densities at the Norcross station at its highest densities. There is quite a bit of scatter, which is indicative of low model precision. Plots of model residual densities versus explanatory variables (figs. 12 to 14) indicate that residuals are distributed fairly equally above and below zero over the full range of the explanatory variables. This indicates that the explanatory variables included in the regression models sufficiently explained the variability in *E. coli* densities and that appropriate mathematical transformations were used to linearize their relations with *E. coli* densities.

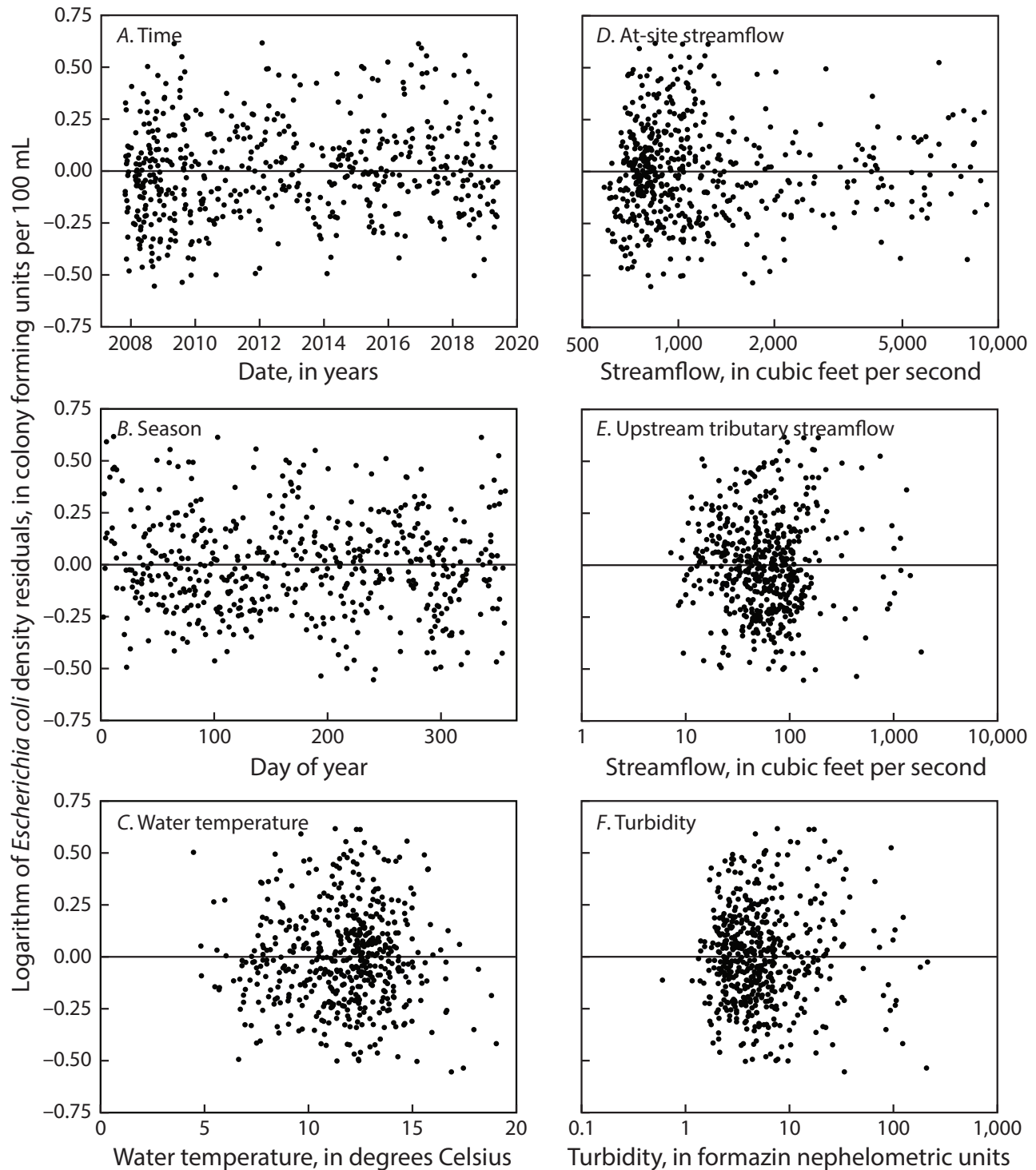


Figure 12. Scatterplots of *Escherichia coli* (*E. coli*) density model residuals versus *A*, time; *B*, season; *C*, water temperature; *D*, at-site streamflow; *E*, upstream tributary streamflow; and *F*, turbidity at Chattahoochee River near Norcross, Ga. (USGS station 02335000) for period October 1, 2007, to May 23, 2019. Black dots are samples used in model calibration ($n = 518$). Logarithms are in base 10. Abbreviation: mL, milliliters.

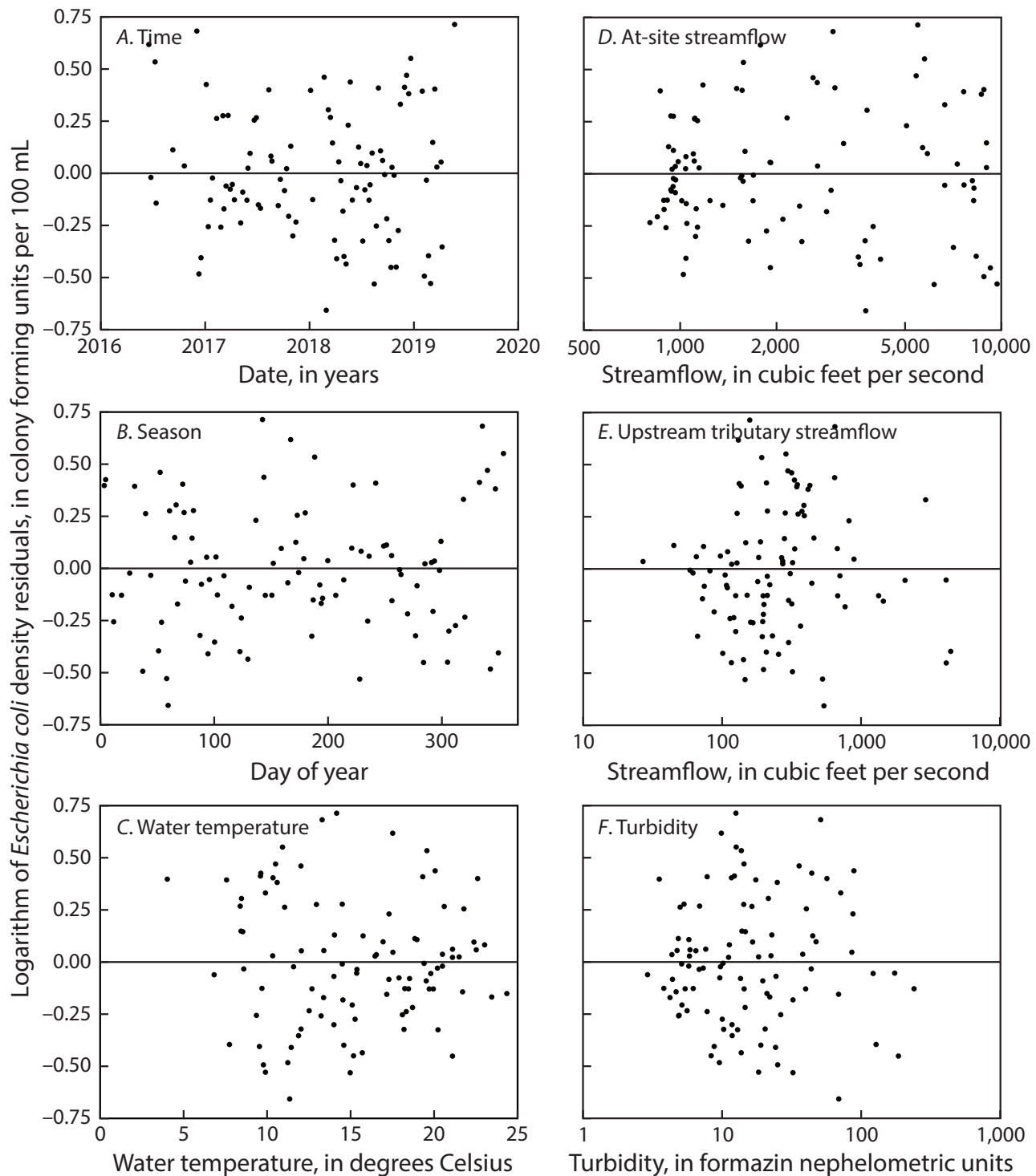


Figure 13. Scatterplots of *Escherichia coli* (*E. coli*) density model versus *A*, time; *B*, season; *C*, water temperature; *D*, at-site streamflow (from USGS station 02336000; lagged by -1 hour); *E*, upstream tributary streamflow; and *F*, turbidity at Chattahoochee River at Powers Ferry and I-285 near Atlanta, Ga. (USGS station 02335880) for period June 16, 2016, to May 23, 2019. Black dots are samples used in model calibration (n = 97). Logarithms are in base 10. Abbreviation: mL, milliliters.

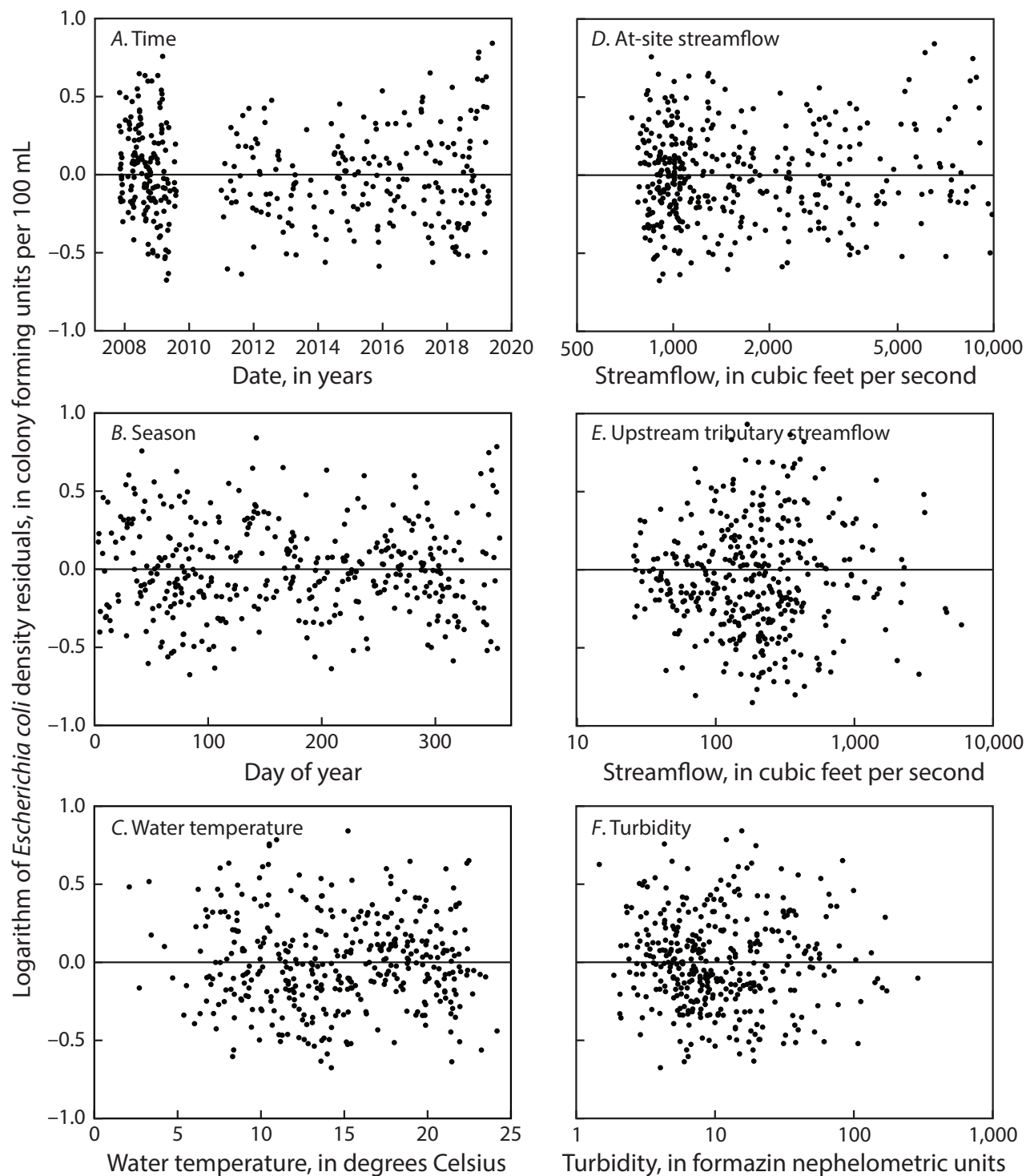


Figure 14. Scatterplots of *Escherichia coli* (*E. coli*) density model residuals versus A, time; B, season; C, water temperature; D, at-site streamflow; E, upstream tributary streamflow; and F, turbidity at Chattahoochee River at Atlanta, Ga. (USGS station 02336000) for period October 4, 2007, to May 23, 2019. Black dots are samples used in model calibration ($n = 380$). Logarithms are in base 10. Abbreviation: mL, milliliters.

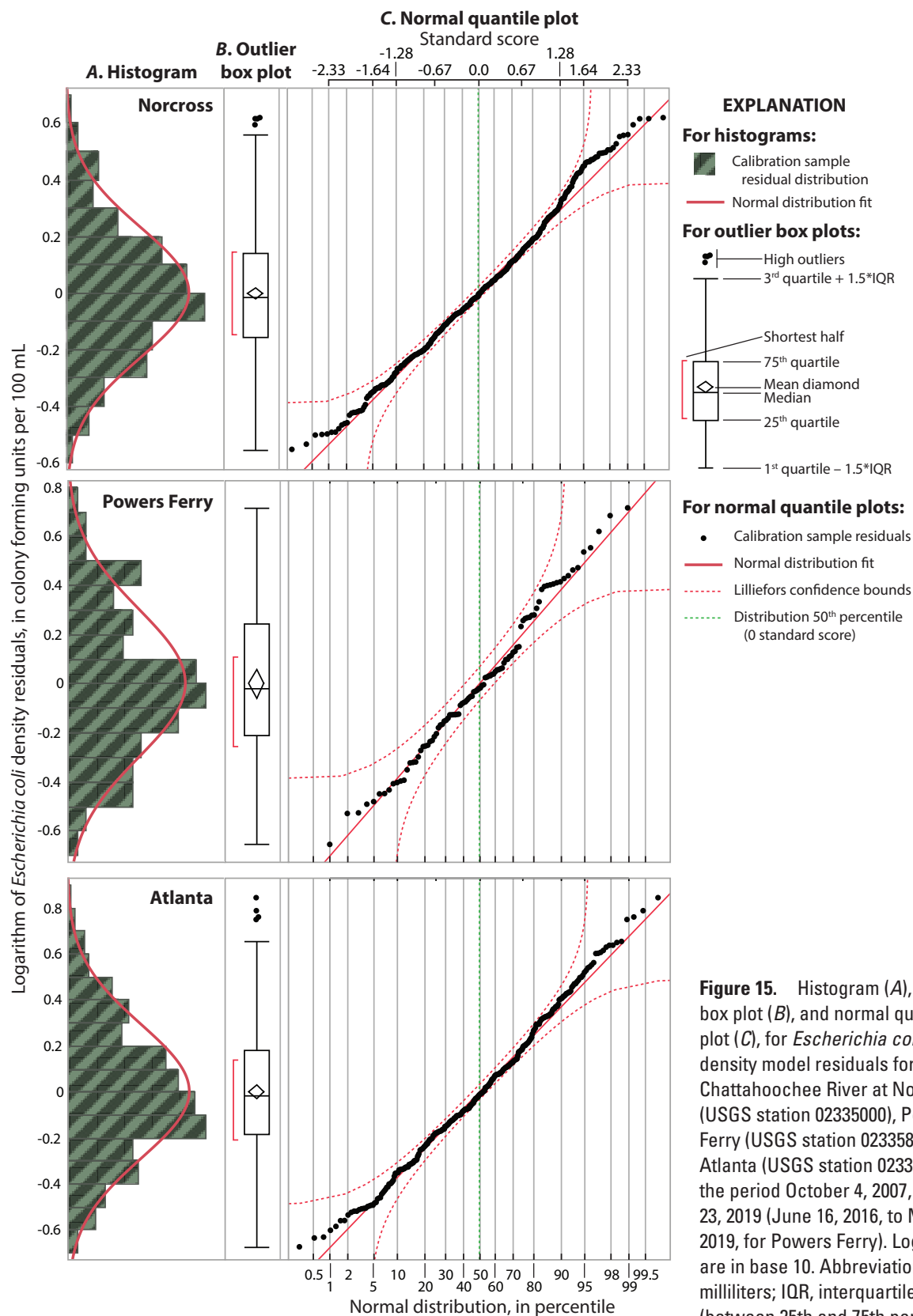


Figure 15. Histogram (A), outlier box plot (B), and normal quantile plot (C), for *Escherichia coli* (*E. coli*) density model residuals for the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 4, 2007, to May 23, 2019 (June 16, 2016, to May 23, 2019, for Powers Ferry). Logarithms are in base 10. Abbreviation: mL, milliliters; IQR, interquartile range (between 25th and 75th percentiles).

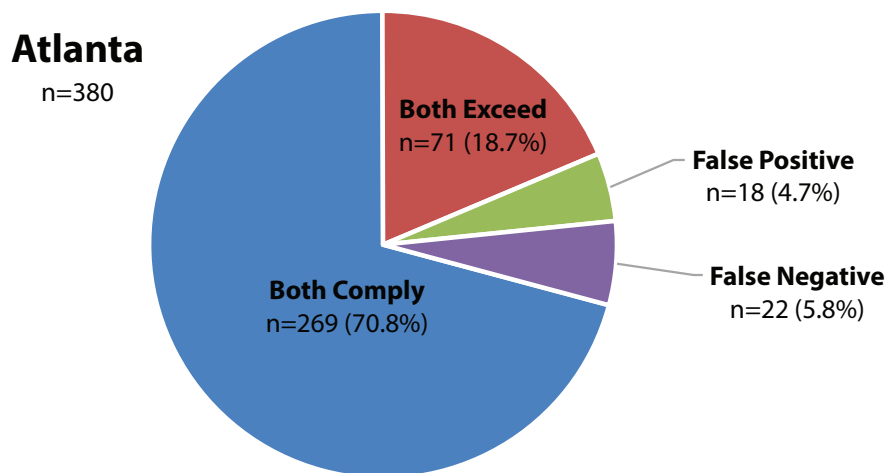
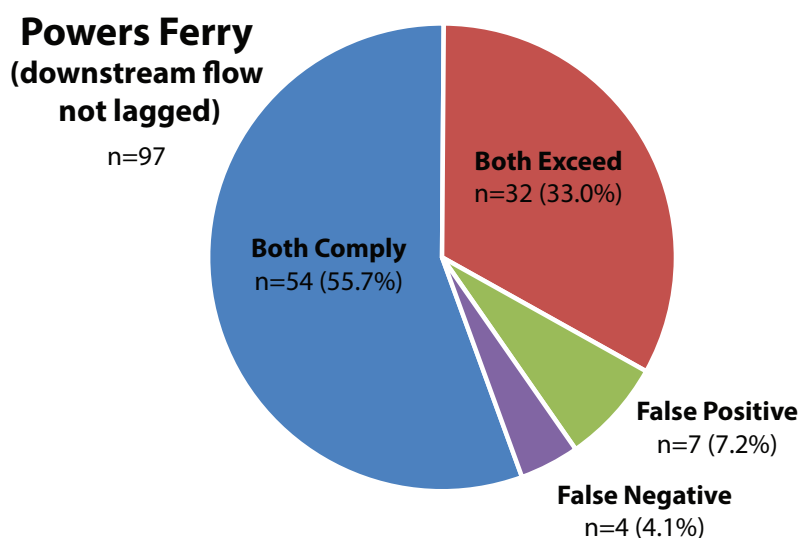
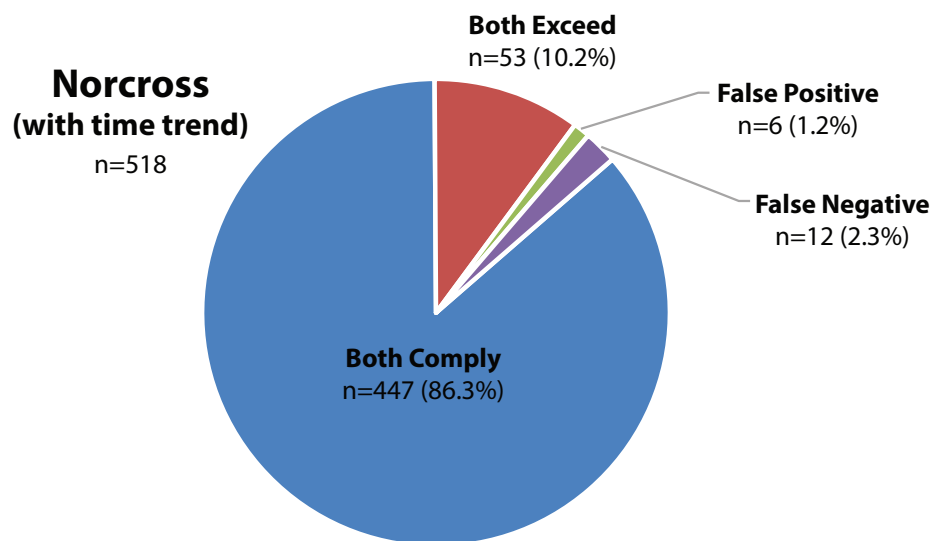


Figure 16. Pie charts showing the frequency observed and model predicted *Escherichia coli* (*E. coli*) densities complied with or exceeded the Beach Action Value (BAV) of 235 colony forming units per 100 mL at the three BacteriALERT stations for the model calibration period (October 4, 2007, to May 23, 2019; June 16, 2016, to May 23, 2019, for Powers Ferry). Does not include outliers. False positives indicate the predicted densities exceeded the BAV while the observed densities did not exceed the BAV; false negatives indicate the observed densities exceeded the BAV while the predicted densities did not exceed the BAV. Abbreviations: mL, milliliters; n, number of samples; %, percent.

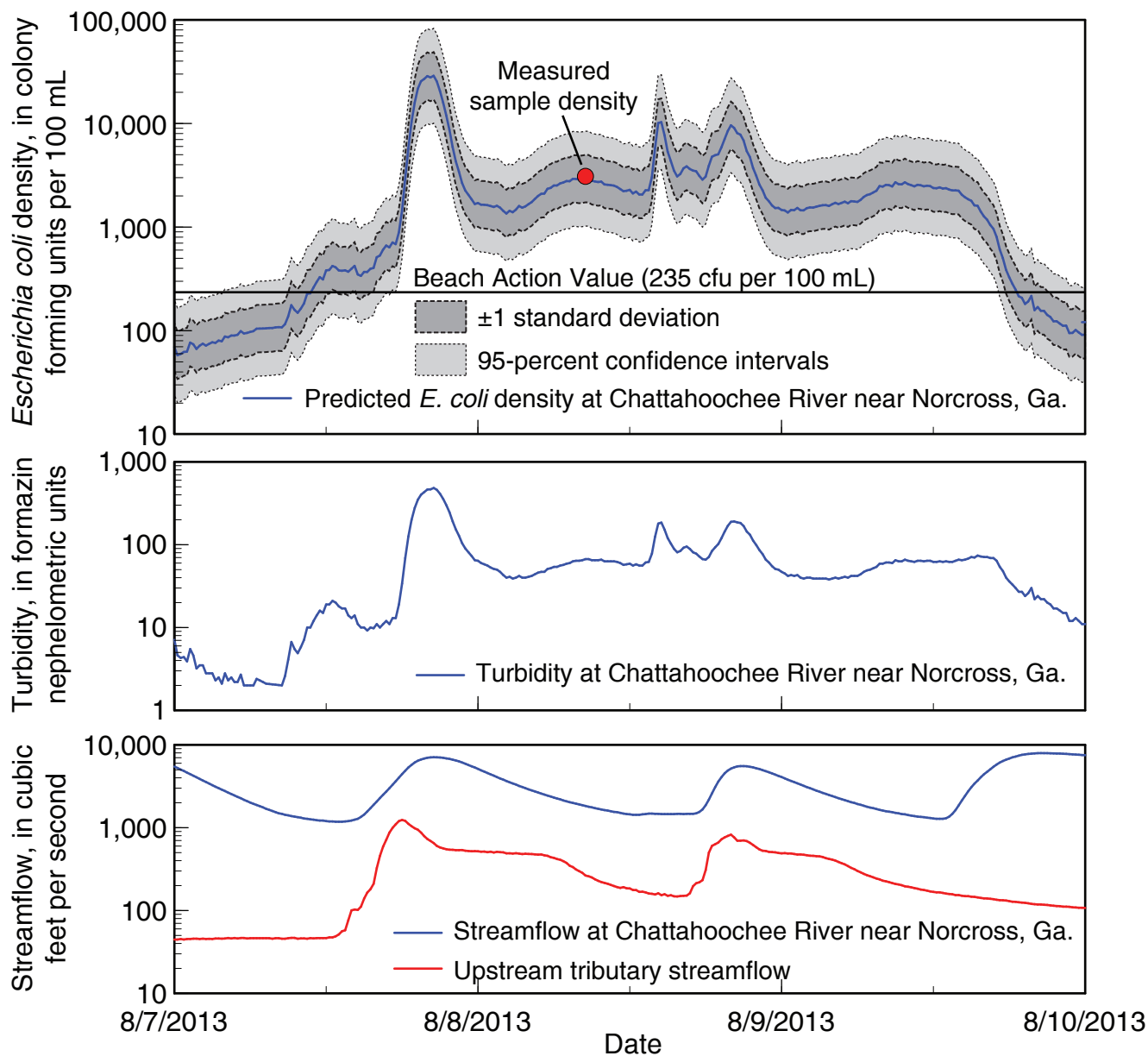


Figure 17. Example of a time-series of *Escherichia coli* (*E. coli*) density model predictions and uncertainty for the Chattahoochee River at Norcross (USGS station 02335000) for the period August 7, 2013, to August 10, 2013. Includes the time-series of the explanatory variables turbidity, at-site streamflow, and upstream tributary streamflow. Abbreviations: cfu, colony forming units; Ga., Georgia; mL, milliliters.

Model residual densities were evaluated for homoscedasticity, normality, and independence. Plots of residuals versus predicted values (fig. 11) and explanatory variables (figs. 12 to 14) indicate that distributions were generally homoscedastic such that variance in error is equal over the entire range of each of these values and variables (even for variables not used in the models). Residual normality test probabilities using the Shapiro-Wilk W Test indicate that residuals for the Powers Ferry models came from a normal distribution (p -value ≥ 0.05) while the residuals from the Norcross and Atlanta models were not from a normal

distribution (table 8). The histograms of residuals fitted with a normal distribution and the residual normal-quantile plots indicate that the residuals were unimodal (single peaked) and reasonably fit a normal distribution with some deviation at the lowest and highest values (fig. 15). This indicates that the distribution of residuals largely do not deviate from normal and that it is unlikely that model fitting and error estimates are substantially affected. Auto correlation of the residuals were low (highest was 0.279 for Atlanta model; table 8) indicating that there was not substantial serial correlation in the residuals that could affect model fit or error estimates.

Table 12. Percentage of time predicted *Escherichia coli* (*E. coli*) densities that exceeded the Beach Action Value (BAV) of 235 colony forming units per 100 mL at the three BacteriALERT stations.

[Abbreviations: cfu, colony forming units; mL, milliliters]

Station/Model	Prediction period	Number of 15-minute interval predictions	Coverage of complete explanatory variables for 15-minute intervals (percent)	Predicted <i>E. coli</i> frequency of exceedance of the BAV (percent)	Model calibration dataset frequency of exceedance of the BAV (percent)
Norcross; with time trend	Oct. 1, 2007, to May 30, 2019	331,856	81.2	14.2	12.5
Powers Ferry; downstream flows not lagged	May 1, 2016, to May 30, 2019	84,852	78.6	34.8	37.1
Atlanta; no time trend	Oct. 1, 2007, to May 30, 2019	247,321	60.5	26.6	24.5

Model Evaluation for Beach Action Value

E. coli model predictions were evaluated to see how well the models performed at predicting exceedance of the BAV criteria of 235 cfu/100 mL; as assessed from corresponding observed sample densities. For the model calibration period (October 4, 2007, to May 23, 2019; June 16, 2016, to May 23, 2019, for Powers Ferry models), the BAV criteria was exceeded 12.5, 37.1, and 24.5 percent of the samples for Norcross, Powers Ferry, and Atlanta stations, respectively (sum of both comply and false negative cases). Misidentification (both false positive and false negative cases) of exceedance of the BAV criteria occurred in 3.5 percent of the samples for Norcross, 11.3 percent of the samples at Powers Ferry, and 10.5 percent of the time at Atlanta (fig. 16). False positives and negatives occurred about equally, as expected since there was no apparent bias in the model predictions. From a health risk perspective, the concern lies with the frequency of false negatives where the model predicted densities below the BAV, but conditions were actually unsafe (observed densities above the BAV). False negatives ranged from 2.3 percent at Norcross to 5.8 percent at Atlanta. The rate of misidentification of exceedance of the criteria is dependent upon the combination of the model precision and how frequently *E. coli* densities are near the BAV. Despite the relatively poor precisions estimated for the predicted densities near the BAV, as indicated by the model RMSEs (table 8), misidentification did not occur frequently because observed densities were not commonly in the range of the BAV (fig. 4).

An example time-series of predicted *E. coli* density and its associated explanatory variables is shown in figure 17 for the Norcross station. Storms occurred on August 7, 2013, and August 8, 2013, as indicated by upstream tributary streamflow. Predicted densities and BAV compliance changed rapidly with changes in turbidity and streamflow. While there were substantial periods where the confidence intervals straddled the BAV, the model uncertainties were more often fully above or below

the BAV. One sample was collected during this timeframe, on August 8, 2013, at 9:18 am, and has a density of 3,100 cfu, which is close to the model predicted value of 2,800 cfu.

The percentage of time the predicted *E. coli* densities exceeded the BAV were determined from the 15-minute interval time series data of the explanatory variables over the time frame of the model calibration period (table 12). The frequency of exceedance of the BAV by the predicted values compared to that of the observed from the calibration samples were somewhat higher for the Norcross and Atlanta models and somewhat lower for the Powers Ferry model. There was also a substantial percentage of time that the time-series explanatory variables were not available for all model parameters preventing predictions. While streamflow is typically estimated to correct for poor data-quality issues, poor turbidity time-series data are removed and hence unavailable. If turbidity data were removed more frequently during particular conditions, it could affect the estimated frequency of exceedance of the BAV. Due to possible lack of representativeness of samples collected and the substantial frequency of missing time-series data, the estimates of exceedance of the BAV should be used only as approximate values for the time period.

Discussion

Evaluation of Model Explanatory Variables

While the regression model approach that is employed herein is empirically based, the variables included in the models allude to some process-based controls on *E. coli* densities. Variables with relatively high logarithmic worth within an individual model is an indication of their larger role in explaining variations in *E. coli* densities, while the other model variables still had a significant, but lesser role. The four most influential variables on predicted densities across

all models, determined by being one of the top three variables (based on logarithm worth) in at least one of the models, included season (day-of-year), upstream tributary streamflow, turbidity, and at-site streamflow (table 10). Of these variables, season and upstream tributary streamflow were included in the top three most influential variables for models at all three stations, while turbidity was in two of the station models, and at-site streamflow was in one of the station models. No single variable stood out as most important across all models as the most influential variable varied by station (turbidity at Norcross, season at Powers Ferry, and upstream tributary streamflow at Atlanta).

The inclusion of upstream tributary streamflow within the top three most influential variables at all station models indicates that this is an important factor in explaining variability in *E. coli* densities and may be a source of high densities that warrant further investigation. This effect is strong enough to override the confounding of Bull Sluice Lake, where mixing attenuates the water-quality signal of the contribution of upstream tributaries for Powers Ferry and Atlanta stations. The Powers Ferry station only includes tributary inflows below the lake from the Sope Creek watershed, and the Atlanta station additionally includes the tributary inflows from the Rottenwood Creek watershed (fig. 3). The effects of mixing in Bull Sluice Lake likely muddles the relations between model explanatory variables and *E. coli* densities and may explain why the Powers Ferry and Atlanta station models have lower R^2 s and higher RMSEs than for the Norcross models. The lack of availability of any one of the tributary flows that comprise this explanatory variable would result in inability to predict *E. coli* densities. Streamflow data are rarely unavailable, but as an alternative, one could prorate other inflows for a missing tributary flow to allow for predictions during these situations.

The lack of inclusion of particular terms is not necessarily an indication of a lack of relation with *E. coli* densities but can be indicative that another variable was (or other variables were) sufficient and better at explaining/accounting for the variability in densities. For example, both high turbidity and high at-site streamflow indicate conditions when suspended sediment concentrations are high, which is associated with high *E. coli* densities. However, turbidity appeared to perform better than at-site streamflow at predicting variations in *E. coli* densities as turbidity was included in all models and had higher logarithmic worth's than at-site streamflow except for the Powers Ferry models. Similarly, only the Atlanta model included a streamwater temperature parameter, possibly indicating that the seasonal terms were sufficient and preferred in modeling the seasonal variability in densities for the other models.

The regression analysis for possible explanatory variables was not fully exhaustive. For example, tributary flows could have been modeled individually, which may have provided insight into which tributaries have the most impact on *E. coli* densities. Other explanatory variables (for example, separate turbidity relations during hydrologic events and an indicator

variable indicating that tributary flows were on the rising limb of the hydrograph) were initially explored but were not found to substantially improve the precision of the predictions.

Assessment of *Escherichia coli* (*E. coli*) Trend Explanatory Variables

The evaluation of trends in *E. coli* density based on the model trend explanatory variable is confounded by the use of the water-quality surrogate-variable turbidity. As long as any changes in *E. coli* density is reflected in turbidity and the relation between *E. coli* density and turbidity remains the same, a model using the turbidity surrogate should accurately capture and predict any variations or trends in *E. coli* density. This is ideal from a predictive modeling standpoint such as is desired for the BacteriALERT program. However, if the interest is in determining whether there are increases or decreases in source contributions of *E. coli*, the use of the water-quality surrogate complicates this evaluation. Furthermore, if changes in source contributions are independent of turbidity, or alters the relation between turbidity and *E. coli* density, a trend could result that would be difficult to evaluate without additional information on how surrogate relations have changed.

The Norcross model has a substantial increasing trend in *E. coli* density with time that is equal to 2.3 percent per year. Two possible explanations for the trend are related to climate patterns and changes in sources of *E. coli* within the CRNRA watershed basin. The climate over the Norcross model calibration period, October 2007 to May 2019, was trending wetter, as substantial droughts occurred more frequently at the beginning and middle of that period (years 2007–2009, 2011, 2012, and 2016). While the at-site and upstream tributary streamflow explanatory variables capture the total streamflow conditions as the result of the climate patterns, wetter conditions typically result in a higher proportion of stormflow relative to base flow that would be associated with higher *E. coli* densities than for the same total streamflow during drier conditions. This mechanism combined with the observed trend toward wetter climate conditions could explain the observed increasing trend in *E. coli* densities at Norcross. However, if climate was the source of trend, a similar trend would be expected to be present in the Atlanta model, but this was not observed; although, the lack of a significant trend in the Atlanta model could be the result of the higher model RMSE that makes it more difficult to detect a trend. A second explanation for the trend at Norcross could be that sources of *E. coli* within the CRNRA watershed basin are changing such that sources have less of a suspended sediment component over time. Turbidity would therefore overestimate *E. coli* densities at the beginning of the model calibration period and underestimate densities at the end of the period, which could explain the increasing trend in *E. coli* densities in the Norcross model while not necessarily indicating that *E. coli* densities are actually increasing over time. The changes in the *E. coli* density versus turbidity

relation could be the result of land use changes (urbanization) and implementation of Best Management Practices (BMPs) such as detention ponds that reduce sediment transport but might not necessarily be as effective at reducing bacteria transport. Both rapid urbanization (Aulenbach and others, 2017a) and implementation of these types of BMPs (Aulenbach and others, 2017b) are known to have occurred in the CRNRA watershed basin upstream of the Norcross station during the model calibration period.

There was an apparent increasing trend in *E. coli* densities at Powers Ferry (fig. 7A). However, the lack of pattern in model residual densities versus decimal year (fig. 13A) indicates that explanatory variables such as at-site and upstream tributary streamflows and turbidity sufficiently captured the variability in *E. coli* densities without needing to employ a trend variable in the model. The apparent trend may be explained by climate patterns over the Powers Ferry model calibration period, May 2016 to May 2019. There was a substantial drought in 2016 while conditions became progressively wetter to the end of the sampling period, which could help explain why observed *E. coli* densities increased over time.

The trend in *E. coli* densities observed in the Norcross model indicates the need to monitor the performance of the *E. coli* model predictions and to update model fits as necessary. It should not be assumed that the trend detected for Norcross will continue in the future and at the same rate. However, the large degree of uncertainty in the model predictions (as indicated by its RMSE) is a limitation in significantly detecting a trend (or change in rate of trend in the case of the Norcross station model). So, it likely will take some time before changes become detectable such that models will not need to be frequently updated.

The USGS has issued policy and guidance for the validation and review of existing surrogate regression models to address future changes in model relations (U.S. Geological Survey, 2016). Ongoing model validation requires continued sampling with steps to be taken when observed densities are greater than two standard errors from model predicted values. Model validation requires an average of eight samples and a minimum of six samples to be collected per year and that at least one sample is collected every quarter, for which the ongoing weekly BacteriALERT sampling fulfills. The USGS requires that models be reviewed annually and refit if necessary, and that models must be refit every three years with the additional validation samples to ensure that models are current and to reduce model uncertainty.

Assessment of Outliers

Outliers were identified based on the magnitudes of their model residual densities and removed to prevent undue influences on the regression model fits. It is possible that outliers were the result of sample contamination or analytical error. It is also possible that there was an infrequent event that resulted

in unusual water quality condition at that time; it was beyond the scope of this analysis to identify events that might explain each outlier. High outliers, which represents 82.8 percent of the total outliers, could be the result of the release of sewage into the Chattahoochee River or its tributaries from sewer leaks and overflows that have been documented in the Atlanta Metropolitan area. Low outliers could occur from the release of primary treated sewage to streams that has undergone emergency chlorination and incomplete dechlorination when treatment capacity is exceeded.

The predictive regression models developed herein are not intended to predict infrequent occurrences that are unrelated to or deviate from the modeled relations with turbidity and streamflow. Removing outliers from the model calibration dataset has the effect of reducing the model RMSE. Hence, excessive removal of outliers results in estimates of model precisions that are better than actually observed and should be avoided. Overall, 3.8 percent of sample densities were identified as outliers and may reflect how frequently model predictions do not represent observed conditions. It may be useful to correlate these outliers (table 1.1, appendix 1) with prior known events to identify conditions where the models do not apply. Additional steps for communicating health risks are warranted during conditions when high outliers reoccur above the BAV, particularly if model predictions indicate compliance with the BAV. When observed sample densities do not match predicted densities that are well outside estimated model precisions, it justifies a reason to investigate for unknown causes.

Model Evaluation for Risk Assessment

The models performed well at predicting whether BAV was exceeded, which was the main objective of developing the predictive models for the BacteriALERT program. Misidentification rate is low, ranging from 3.5 percent of the time for Norcross to 11.3 percent of the time at Powers Ferry. The frequency of false negatives, in which model predictions indicate safe conditions while observed densities indicate that the BAV criteria was actually exceeded, ranged from 2.3 percent at Norcross to 5.8 percent at Atlanta. The predictive *E. coli* models provide a substantial benefit in communicating the health risk in near real-time in comparison to assessing the health risk from observed sample densities from samples collected once per week with results delayed 18-hours for analysis. This is especially pertinent during storm conditions when *E. coli* densities often change rapidly.

While the model predicted *E. coli* densities were sufficient for indicating exceedance of the BAV, they were not particularly precise, as indicated by their model RMSEs (table 11). This is not too surprising as similar errors have been reported for estimating suspended sediment to which a substantial portion of bacteria can be attached and likewise can have a lot of natural variability due to inconsistent variations in transport processes. The calculated precision of model predictions represents a combination of (1) how

Table 13. Comparison of *Escherichia coli* (*E. coli*) model statistics between models developed in this report and in Lawrence (2012).

[Abbreviation: na, not applicable]

Station	Model	Adjusted model R ²	Root-mean-square error (RMSE)	Source in Lawrence (2012)
Norcross	This report, with trend term	0.795	0.229	na
	Log turbidity	0.512	0.394	Table 8, model 1
	Log turbidity, event, season, hydrologic condition, event x log turbidity	0.791	0.242	Table 8, model 8
Atlanta	This report, no trend term	0.686	0.292	na
	Log turbidity	0.496	0.408	Table 13, model 9
	Log turbidity, water temperature, event, event x log turbidity	0.758	0.254	Table 13, model 12

well explanatory variables can model variations in densities, (2) variability related to sampling, and (3) analytical accuracy. Lawrence (2012), based on an analysis of replicates using data from the Norcross and Atlanta stations for the period 2000 to 2008, reported that analytical precisions based on replicates from sample dilutions ranged from 1.3 to 17 percent and the relative percent difference (RPD) of sample duplicates ranged from 3.6 to 35 percent. Hence, much of the poor model precision might be attributed to variability in densities that was not explained by the models. *E. coli* densities and suspended sediment concentrations typically vary over several orders of magnitude, and while the model R²s of the bacteria models indicate they explain a large portion of the variability in densities, a large amount of variability remains in the model residuals.

The model decision criteria could be adjusted lower than the BAV criteria to reduce the frequency of false negatives and account for the additional uncertainty associated with using model predictions. The BAVs (table 1) were determined by the Environmental Protection Agency (EPA) using observed samples which have lower uncertainty than the model predicted densities. However, there was no substantial bias in the predicted values and the misidentification rate was low; therefore, on average, the predicted values reflect the acceptable illness rate corresponding to the BAV criteria with infrequent errors. Lowering the model decision criteria would result in more frequent false positives where predictions indicate unsafe conditions while conditions actually comply with the BAV. This would effectively lower what is considered the acceptable illness rate.

The BacteriALERT *E. coli* models predict bacteria conditions at the locations of the stations. However, water-based recreation occurs throughout the CRNRA such that the health risk needs to be approximated for adjacent and intervening river reaches. This is particularly important for recreational activities where one travels along the river, such as tubing, canoeing, kayaking, boating, and rowing. Adding the BacteriALERT station at Powers Ferry improved the resolution of bacteria conditions in this downstream reach that is heavily

used for recreation. *E. coli* densities at the Powers Ferry and Atlanta stations were generally higher and varied more similarly than at the Norcross station (fig. 5). The Powers Ferry and Atlanta stations more frequently exceeded the BAV than the Norcross station (table 6). If bacteria levels comply with the BAV criteria at both the upstream and downstream BacteriALERT stations, this would indicate bacteria levels would likely be similar along this reach of the Chattahoochee River. However, if conditions exceeded the BAV at either one of the upstream or downstream stations, it is difficult to say with confidence where conditions become unsafe as one moves further away from the vicinity of the station that complies with the BAV criteria. Additional sampling along the Chattahoochee River would be needed to better determine temporal and longitudinal patterns relative to the three BacteriALERT stations' *E. coli* density variations.

Comparison with Previous Models

Lawrence (2012) did extensive modeling of *E. coli* densities at the Norcross and Atlanta stations. Lawrence fit a series of models with increasing complexity for each station. The models developed herein were compared to the simple turbidity-only Lawrence models, which are currently (as of 2020) implemented in the BacteriALERT program for alerting recreational users, and the more complex Lawrence models that have the lowest RMSEs (table 13). The models developed herein outperform the simple turbidity models, as the models in this report have higher model R²s, and more importantly, substantially lower RMSEs (table 13). The Norcross model in this report had a somewhat smaller RMSE than the complex Lawrence (2012) model while the Atlanta model had a larger RMSE than the complex model in Lawrence. This indicates that the model from this report is somewhat more precise for the Norcross station, while the Lawrence model is more precise for the Atlanta station. However, model RMSEs are not exactly comparable between this report and from Lawrence (2012). While the models in both reports were

developed from data from the same BacteriALERT program and stations, the models were calibrated from samples collected from different periods with only a short overlap; October 2000 to September 2008 for the Lawrence models and October 4, 2007, to May 23, 2019, herein. Relations may be different during these periods and the difficulty in fitting the variation in *E. coli* densities might not be equal due to idiosyncrasies in the calibration datasets. In both analyses, outliers were removed through the use of different criteria, which can have a substantial effect on the estimated RMSE.

The BacteriALERT program currently uses simple log turbidity models, despite better models being available that have lower RMSEs, as to avoid the difficulties in determining the various indicator variables used in the more computationally complex Lawrence models. Implementation of the models developed herein will be more readily implemented due to the use of variables assessable from the USGS time-series database. Lawrence did not provide a retransformation bias correction factor to account for the bias in model predictions due to modeling in logarithmic space and making estimations in arithmetic space. Correction factors for the models developed herein indicate that models would underestimate *E. coli* densities by 13.3 to 19.7 percent without these correction factors applied.

The more complex Lawrence models used similar explanatory variables to the models developed herein. All models relied on turbidity as one of the most important variables. For both the models in this report and in Lawrence (2012), water temperature was similarly included in Atlanta station models and excluded from the Norcross station models. Models in this report used upstream tributary streamflows as explanatory variables while Lawrence utilized event and hydrologic condition indicator variables to capture when there were stormflow contributions from its tributaries.

Summary and Conclusions

In 2000, the National Park Service entered into a public-private partnership with the U.S. Geological Survey (USGS) and the Chattahoochee Riverkeeper, called the Chattahoochee River BacteriALERT program, to monitor *E. coli* densities within the Chattahoochee River National Recreation Area (CRNRA). The CRNRA (<https://www.nps.gov/chat/index.htm>) is a National Park Service unit/park with 48 miles of urban waterway in the Atlanta metropolitan area. *E. coli* is a proxy for human health risk to water-based recreators from waterborne pathogens, which can result in gastrointestinal illnesses. The BacteriALERT currently (as of 2020) monitors *E. coli* densities at three stations on the Chattahoochee River within the CRNRA, at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880; sampling started in May 2016), and Atlanta (USGS station 02336000; <https://www2.usgs.gov/water/southatlantic/ga/bacteria/index.php>). In addition, *E. coli* densities have been

predicted in near real-time at the Norcross and Atlanta stations since approximately 2012 using empirical models to provide more timely notification of health risks (Lawrence, 2012). This report documents monitoring results since 2000, the development of updated regression models for predicting *E. coli* densities at the Norcross and Atlanta stations, and new models developed for the Powers Ferry station, which was installed in 2016.

Over the period of the study, *E. coli* samples have been collected at various sampling frequencies and are currently collected weekly. Samples were analyzed for *E. coli* using standard method 9223B with the Colilert®-18 medium and Quanti-Tray® system. *E. coli* densities were compared to the Environmental Protection Agency's (EPAs) Beach Action Value (BAV) of 235 colony forming units per 100 milliliters (235 cfu/mL) to assess whether conditions were considered safe for recreational use. The BAV criteria corresponds to EPA's National Epidemiological and Environmental Assessment of Recreational Water gastrointestinal illness rate of 36 per 1,000 primary contact recreators. Sample *E. coli* densities exceeded the BAV 15.5 percent of the time at Norcross ($n = 1,969$) and 30.3 percent of the time at Atlanta ($n = 1,938$) for the period between October 23, 2000, to May 23, 2019, and for 33.6 percent of the time at Powers Ferry ($n = 134$) for the period between May 5, 2016, to May 23, 2019.

E. coli densities were predicted in near real-time from models developed using forward-stepwise, multiple linear regression with the Bayesian Information Criteria and were calibrated from samples collected between October 4, 2007, and May 23, 2019. Explanatory variables included season, turbidity, water temperature, at-site and upstream tributary streamflows, and time trend. The most important explanatory variables among models were turbidity, upstream tributary streamflows, and season. The inclusion of upstream tributary streamflows in all the models indicates that tributaries in the CRNRA watershed basin may be an important source of *E. coli* and could be a good focus for future mitigation. The Norcross model required a trend term to capture the 2.3 percent increase in *E. coli* densities per year over the calibration period. Outliers were identified and removed from the model calibration datasets and represent 3.8 percent of sample densities, which may reflect how frequently model predictions do not represent observed conditions. Most outliers had high densities and may have reflected periods of release of sewage into the Chattahoochee River or its tributaries from sewer overflows, conveyance leaks, or from perturbations at sewage treatment plants. The models explained a substantial portion of the variations in *E. coli* densities, with model adjusted R^2 s ranging from 0.686 (Atlanta) to 0.795 (Norcross with time trend). Evaluation of model predictions and residuals indicated that models were well posed and exhibited little bias. The models performed well in accurately determining compliance and exceedance of the BAV with low misidentification rates ranging from 3.5 percent (Norcross) to 11.3 percent (Powers Ferry) and with little bias, as indicated by near equivalent

cases of false positives and false negatives. The rate of misidentification was dependent upon the combination of model precision and how frequently *E. coli* densities were near the BAV. Actual model predictions were not overly precise as indicated by their model RMSEs (0.229 to 0.310), which is typical for sediment related constituents.

The models developed herein are an improvement on the simple turbidity only models currently (as of 2020) employed by the BacteriALERT program. The currently employed models have higher RMSE compared to the models developed herein and were not corrected for back-transformation bias. The RMSEs of the models developed herein were comparable to the more complex models developed by Lawrence (2012) but are more readily implemented due to the lack of indicator variables that were computationally difficult to determine. *E. coli* densities can change rapidly during storm conditions, and assessments from observed sample densities from samples collected once per week (with results delayed 24 hours for analysis) are not timely. The predictive *E. coli* models developed herein should provide a substantial improvement in assessing potential health risks of water-based recreational activities in the CRNRA in near real-time.

References Cited

- Akaike, H., 1974, A new look at the statistical model identification: IEEE Transactions on Automatic Control, v. 19, no. 6, p. 716–723, accessed April 20, 2020, at <https://doi.org/10.1109/TAC.1974.1100705>.
- Anderson, C.W., 2005, Turbidity (ver. 2.1, September 2005): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A6.7, 53 p., accessed April 20, 2020, at <https://doi.org/10.3133/twri09A6.7>.
- American Society for Testing and Materials International, 2003, D1889–00 Standard test method for turbidity of water, in ASTM International, Annual Book of ASTM Standards [Water and Environmental Technology]: v. 11.01, sec. 11, West Conshohocken, Pennsylvania, 6 p.
- Aulenbach, B.T., Joiner, J.K., and Painter, J.A., 2017a, Hydrology and water quality in 13 watersheds in Gwinnett County, Georgia, 2001–15: U.S. Geological Survey Scientific Investigations Report 2017–5012, 82 p., accessed April 20, 2020, at <https://doi.org/10.3133/sir20175012>.
- Aulenbach, B.T., Landers, M.N., Musser, J.W., and Painter, J.A., 2017b, Effects of impervious area and BMP implementation and design on storm runoff and water quality in eight small watersheds: Journal of the American Water Resources Association, v. 53, no. 2, p. 382–399, accessed April 20, 2020, at <https://doi.org/10.1111/1752-1688.12501>.
- Booth, E.A., and Adams, V.A., 2018, Bacteria criteria for drinking water and fishing designated uses—Technical support document for the proposed criteria to protect secondary recreators: Georgia Department of Natural Resources, Georgia Environmental Protection Division, Atlanta, Georgia, 21 p., accessed April 20, 2020, at <https://epd.georgia.gov/document/publication/tsd-bacteria-criteria-finalpdf/download>.
- Clesceri, L.S., Greenberg, A.E., and Eaton, A.D., eds., 1998, Standard methods for the examination of water and wastewater (20th ed.): Washington, D.C., American Public Health Association-American Water Works Association-Water Environment Federation, p. 9.21, 9.63–65, 9.69–70.
- Cullinane, T.C., Koontz, L., and Cornachione, E., 2019, 2018 National Park visitor spending effects—Economic contributions to local communities, States, and the Nation: Natural Resource Report NPS/NRSS/EQD/NRR—2019/1922, Fort Collins, Colorado, National Park Service, 55 p., accessed April 20, 2020, at https://www.nps.gov/nature/customcf/NPS_Data_Visualization/docs/NPS_2018_Visitor_Spending_Effects.pdf.
- Darakas, E., 2002, *E. coli* kinetics—Effect of temperature on the maintenance and respectively the decay phase: Environmental Monitoring and Assessment, v. 78, no. 2, p. 101–110, accessed April 20, 2020, at <https://doi.org/10.1023/A:1016313520396>.
- Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, no. 383, p. 605–610, accessed April 20, 2020, at <https://doi.org/10.1080/01621459.1983.10478017>.
- Ferguson, R.I., 1986, River loads underestimated by rating curves: Water Resources Research, v. 22, no. 1, p. 74–76, accessed April 20, 2020, at <https://doi.org/10.1029/WR022i001p00074>.
- Fries, J.S., Characklis, G.W., and Noble, R.T., 2006, Attachment of fecal indicator bacteria to particles in the Neuse River Estuary, North Carolina: Journal of Environmental Engineering, v. 132, no. 10, p. 1338–1345, accessed April 20, 2020, at [https://doi.org/10.1061/\(ASCE\)0733-9372\(2006\)132:10\(1338\)](https://doi.org/10.1061/(ASCE)0733-9372(2006)132:10(1338)).
- Gregory, M.B., and Frick, E.A., 2000, Fecal-coliform bacteria concentrations in streams of the Chattahoochee River National Recreation Area, Metropolitan Atlanta, Georgia, May–October 1994 and 1995: U.S. Geological Survey Water-Resources Investigations Report 2000–4139, 8 p., accessed April 20, 2020, at <https://doi.org/10.3133/wri20004139>.

- Hartel, P.G., Frick, E.A., Funk, A.L., Hill, J.L., Summer, J.D., and Gregory, M.B., 2004, Sharing of ribotype patterns of *Escherichia coli* isolates during baseflow and stormflow conditions: U.S. Geological Survey Scientific Investigations Report 2004–5004, 10 p., accessed April 20, 2020, at <https://doi.org/10.3133/sir20045004>.
- Helsel, D.R., and Hirsch, R.M., 1992, Statistical Methods in Water Resources: Amsterdam, The Netherlands, Elsevier Science B.V, 529 p.
- IDEXX Laboratories, Inc., 2002a, Colilert®-18 test kit: accessed February 5, 2003, at <https://www.idexx.com/water/refs/060202711C18.pdf>.
- IDEXX Laboratories, Inc., 2002b, Quanti-Tray/2000®: accessed February 5, 2003, at <https://www.idexx.com/water/refs/0602320.pdf>.
- JMP®, 1989–2019, ver. 14.3.0: Cary, North Carolina, SAS Institute Inc.
- Krometis, L.-A.H., Characklis, G.W., Simmons, O.D., III, Dilts, M.J., Likirdopulos, C.A., and Sobsey, M.D., 2007, Intra-storm variability in microbial partitioning and microbial loading rates: Water Research, v. 41, no. 2, p. 506–516, accessed April 20, 2020, at <https://doi.org/10.1016/j.watres.2006.09.029>.
- Lawrence, S.J., 2012, *Escherichia coli* bacteria density in relation to turbidity, streamflow characteristics, and season in the Chattahoochee River near Atlanta, Georgia, October 2000 through September 2008—Description, statistical analysis, and predictive modeling: U.S. Geological Survey Scientific Investigations Report 2012–5037, 81 p., accessed April 20, 2020, at <https://doi.org/10.3133/sir20125037>.
- Leclerc, H., Mossel, D., Edberg, S., and Struijk, C., 2001, Advances in the bacteriology of the coliform group—Their suitability as markers of microbial water safety: Annual Review of Microbiology, v. 55, no. 1, p. 201–234, accessed April 20, 2020, at <https://doi.org/10.1146/annurev.micro.55.1.201>.
- McCulloch, C.E., and Neuhaus, J.M., 2011, Misspecifying the shape of a random effects distribution—Why getting it wrong may not matter: Statistical Science, v. 26, no. 3, p. 388–402, accessed April 20, 2020, at <https://doi.org/10.1214/11-STS361>.
- McKee, A., Molina, M., Cyterski, M., and Couch, A., 2020, Microbial source tracking (MST) in Chattahoochee River National Recreation Area—Seasonal and precipitation trends in MST marker concentrations, and associations with *E. coli* levels, pathogenic marker presence, and land use: Water Research, v. 171, 12 p., accessed April 20, 2020, at <https://doi.org/10.1016/j.watres.2019.115435>.
- McLellan, S.L., Daniels, A.D., and Salmore, A.K., 2001, Clonal populations of thermotolerant *Enterobacteriaceae* in recreational water and their potential interference with fecal *Escherichia coli* counts: Applied and Environmental Microbiology, v. 67, no. 10, p. 4934–4938, accessed April 20, 2020, at <https://doi.org/10.1128/AEM.67.10.4934-4938.2001>.
- National Park Service, 2011, Chattahoochee River National Recreation Area—Long-range interpretive plan: 58 p., accessed January 2015, at https://www.nps.gov/hfc/pdf/ip/CHAT_LRIP.pdf.
- Rantz, S.E., and others, 1982a, Measurement and computation streamflow—Volume 1, measurement of stage and discharge: U.S. Geological Survey Water-Supply Paper 2175, 313 p., accessed April 20, 2020, at <https://doi.org/10.3133/wsp2175>.
- Rantz, S.E., and others, 1982b, Measurement and computation of streamflow—Volume 2, computation of discharge: U.S. Geological Survey Water-Supply Paper 2175, 373 p., accessed April 20, 2020, at <https://doi.org/10.3133/wsp2175>.
- Rasmussen, P.P., Gray, J.R., Glysson, G.D., and Ziegler, A.C., 2009, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods, book 3, chap. C4, 52 p., accessed April 20, 2020, at <https://doi.org/10.3133/tm3C4>.
- Riedel, T.E., Thulsiraj, V., Zimmer-Faust, A.G., Dagit, R., Krug, J., Hanley, K.T., Adamek, K., Ebentier, D.L., Torres, R., Cobian, U., Peterson, S., and Jay, J.A., 2015, Long-term monitoring of molecular markers can distinguish different seasonal patterns of fecal indicating bacteria sources: Water Research, v. 71, p. 227–243, accessed April 20, 2020, at <https://doi.org/10.1016/j.watres.2014.12.037>.
- Schoen, M.E., and Ashbolt, N.J., 2010, Assessing pathogen risk to swimmers at non-sewage impacted recreational beaches: Environmental Science & Technology, v. 44, no. 7, p. 2286–2291, accessed April 20, 2020, at <https://doi.org/10.1021/es903523q>.
- Schwarz, G.E., 1978, Estimating the dimension of a model: Annals of Statistics, v. 6, no. 2, p. 461–464, accessed April 20, 2020, at <https://doi.org/10.1214/aos/1176344136>.
- Seeger, M., Errea, M.-P., Beguería, S., Arnáez, J., Martí, C., and García-Ruiz, J.M., 2004, Catchment soil moisture and rainfall characteristics as determinant factors for discharge/suspended sediment hysteretic loops in a small headwater catchment in the Spanish Pyrenees: Journal of Hydrology (Amsterdam), v. 288, nos. 3–4, p. 299–311, accessed April 20, 2020, at <https://doi.org/10.1016/j.jhydrol.2003.10.012>.

- Soller, J.A., Schoen, M.E., Bartrand, T., Ravenscroft, J.E., and Ashbolt, N.J., 2010, Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination: *Water Research*, v. 44, no. 16, p. 4674–4691, accessed April 20, 2020, at <https://doi.org/10.1016/j.watres.2010.06.049>.
- Tallon, P., Magajna, B., Lofranco, C., and Leung, K.T., 2005, Microbial indicators of faecal contamination in water—A current perspective: *Water, Air, and Soil Pollution*, v. 166, no. 1–4, p. 139–166, accessed April 20, 2020, at <https://doi.org/10.1007/s11270-005-7905-4>.
- U.S. Environmental Protection Agency, 2003, Guidelines establishing test procedures for the analysis of pollutants; Analytical methods for biological pollutants in ambient water; Final rule: *Federal Register*, part 3, 40 CFR Part 136, v. 68, no. 139, p. 43272–43283, assessed April 20, 2020, at <https://www.govinfo.gov/content/pkg/FR-2003-07-21/pdf/03-18155.pdf>.
- U.S. Environmental Protection Agency, 2012, *Recreational Water Quality Criteria*: U.S. Environmental Protection Agency Office of Water report number 820-F-12-058, 63 p., assessed April 20, 2020, at <https://www.epa.gov/sites/production/files/2015-10/documents/rwqc2012.pdf>.
- U.S. Environmental Protection Agency, 2016, Rule 391-3-6.03—Water use classifications and water quality standards [Georgia, 2016]: Georgia 2016, 95 p., accessed April 20, 2020, at <https://www.epa.gov/sites/production/files/2014-12/documents/gawqs.pdf>.
- U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, 40 p., accessed April 20, 2020, at <https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf>.
- U.S. Geological Survey, various dates, *National field manual for the collection of water-quality data*: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A10, accessed April 20, 2020, at <https://pubs.water.usgs.gov/twri9A>.
- Wade, T.J., Sams, E.A., Haugland, R., Brenner, K.P., Li, Q., Wymer, L., Molina, M., Oshima, K., and Dufour, A.P., [2010], Report on 2009 National epidemiologic and environmental assessment of recreational water epidemiology studies: U.S. Environmental Protection Agency Report EPA/600/R-10/168, 450 p., accessed April 20, 2020, at https://archive.epa.gov/near/web/pdf/report2009v5_508comp.pdf.
- Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A., 2006, Guidelines and standard procedures for continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1–D3, 51 p., 8 attachments, accessed April 20, 2020, at <https://doi.org/10.3133/tm1D3>.

Appendix 1. Outliers Removed from Regression Analysis

Table 1.1 contains a list of outliers identified and removed from the regression models for the model calibration period October 4, 2007, to May 23, 2019. Overall, 3.8 percent of sample densities were removed from the calibration dataset (table 7).

Table 1.1. Outlier *Escherichia coli* (*E. coli*) densities removed from regression analysis for the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 4, 2007, to May 23, 2019.

[Predicted *E. coli* densities from regression models selected for use in BacteriALERT program. Last column shows compliance of the predicted and observed *E. coli* densities with the Beach Action Value (BAV) of 235 colony forming units per 100 mL; a false positive indicates the predicted density exceeded the BAV while the observed density did not exceed the BAV; a false negative indicates the observed density exceeded the BAV while the predicted density did not exceed the BAV. Abbreviations: EST, Eastern Standard Time; MPN/100 mL, most probable number per 100 milliliters; USGS, U.S. Geological Survey]

Outlier number	USGS station location	Sample Date, time (EST)	Observed <i>E. coli</i> density outlier (MPN/100 mL)	Outlier type	Predicted <i>E. coli</i> density (MPN/100 mL)	Compliance with BAV
1	Norcross	Dec. 20, 2007, 11:40 a.m.	300	high	32	False negative
2	Norcross	July 31, 2008, 11:20 a.m.	49	low	222	Both comply
3	Norcross	Aug. 27, 2008, 10:15 a.m.	4,900	low	31,069	Both exceed
4	Norcross	Sept. 24, 2008, 10:37 a.m.	12	low	53	Both comply
5	Norcross	Dec. 10, 2008, 11:32 a.m.	1,100	high	155	False negative
6	Norcross	Jan. 15, 2009, 11:31 a.m.	300	high	35	False negative
7	Norcross	Feb. 18, 2009, 10:19 a.m.	260	high	27	False negative
8	Norcross	Mar. 16, 2009, 8:45 a.m.	870	high	217	False negative
9	Norcross	Mar. 26, 2009, 10:00 a.m.	260	high	34	False negative
10	Norcross	Sept. 14, 2009, 9:10 a.m.	290	high	67	False negative
11	Norcross	June 2, 2011, 10:16 a.m.	290	high	58	False negative
12	Norcross	Apr. 19, 2012, 10:32 a.m.	840	high	126	False negative
13	Norcross	Aug. 2, 2012, 11:46 a.m.	560	high	115	False negative
14	Norcross	July 11, 2013, 8:58 a.m.	650	high	71	False negative
15	Norcross	Aug. 15, 2013, 8:22 a.m.	520	high	130	False negative
16	Norcross	Sept. 5, 2014, 9:51 a.m.	5,400	high	914	Both exceed
17	Norcross	Feb. 25, 2016, 10:47 a.m.	310	high	59	False negative
18	Norcross	Feb. 14, 2019, 11:17 a.m.	62	high	14	Both comply
19	Powers Ferry	June 15, 2017, 8:30 a.m.	1,500	high	348	Both exceed
20	Powers Ferry	Aug. 31, 2017, 8:46 a.m.	1,400	high	308	Both exceed
21	Powers Ferry	June 7, 2018, 8:07 a.m.	60	low	341	False positive
22	Powers Ferry	Apr. 18, 2019, 8:55 a.m.	30	low	191	Both comply
23	Powers Ferry	Apr. 25, 2019, 8:56 a.m.	2,100	high	248	Both exceed
24	Powers Ferry	May 2, 2019, 9:50 a.m.	1,400	high	168	False negative
25	Powers Ferry	May 9, 2019, 8:47 a.m.	2,400	high	169	False negative
26	Powers Ferry	May 16, 2019, 9:03 a.m.	2,400	high	203	False negative
27	Atlanta	Oct. 11, 2007, 8:50 a.m.	960	high	149	False negative
28	Atlanta	Jan. 23, 2008, 9:15 a.m.	480	high	76	False negative
29	Atlanta	Mar. 31, 2008, 7:40 a.m.	1,100	high	108	False negative
30	Atlanta	July 7, 2008, 8:35 a.m.	2,100	high	290	Both exceed
31	Atlanta	Feb. 12, 2009, 8:50 a.m.	350	high	65	False negative

Table 1.1. Outlier *Escherichia coli* (*E. coli*) densities removed from regression analysis for the Chattahoochee River at Norcross (USGS station 02335000), Powers Ferry (USGS station 02335880), and Atlanta (USGS station 02336000) for the period October 4, 2007, to May 23, 2019.—Continued

[Predicted *E. coli* densities from regression models selected for use in BacteriALERT program. Last column shows compliance of the predicted and observed *E. coli* densities with the Beach Action Value (BAV) of 235 colony forming units per 100 mL; a false positive indicates the predicted density exceeded the BAV while the observed density did not exceed the BAV; a false negative indicates the observed density exceeded the BAV while the predicted density did not exceed the BAV. Abbreviations: EST, Eastern Standard Time; MPN/100 mL, most probable number per 100 milliliters; USGS, U.S. Geological Survey]

Outlier number	USGS station location	Sample Date, time (EST)	Observed <i>E. coli</i> density outlier (MPN/100 mL)	Outlier type	Predicted <i>E. coli</i> density (MPN/100 mL)	Compliance with BAV
32	Atlanta	Mar. 1, 2012, 9:31 a.m.	920	high	190	False negative
33	Atlanta	Feb. 27, 2014, 8:15 a.m.	12	low	115	Both comply
34	Atlanta	Apr. 14, 2016, 8:03 a.m.	32	low	226	Both comply
35	Atlanta	Dec. 1, 2016, 10:10 a.m.	3,500	high	563	Both exceed
36	Atlanta	Aug. 31, 2017, 9:16 a.m.	4,000	high	471	Both exceed
37	Atlanta	Apr. 25, 2019, 9:15 a.m.	2,300	high	156	False negative
38	Atlanta	May 2, 2019, 10:15 a.m.	1,800	high	115	False negative
39	Atlanta	May 9, 2019, 9:27 a.m.	2,800	high	132	False negative
40	Atlanta	May 16, 2019, 9:22 a.m.	2,400	high	151	False negative

